Microdiversification in genome-streamlined ubiquitous freshwater Actinobacteria

Neuenschwander, Stefan M ; Ghai, Rohit ; Pernthaler, Jakob ; Salcher, Michaela M

Abstract: Actinobacteria of the acI lineage are the most abundant microbes in freshwater systems, but there are so far no pure living cultures of these organisms, possibly because of metabolic dependencies on other microbes. This, in turn, has hampered an in-depth assessment of the genomic basis for their success in the environment. Here we present genomes from 16 axenic cultures of acI Actinobacteria. The isolates were not only of minute cell size, but also among the most streamlined free-living microbes, with extremely small genome sizes (1.2–1.4 Mbp) and low genomic GC content. Genome reduction in these bacteria might have led to auxotrophy for various vitamins, amino acids and reduced sulphur sources, thus creating dependencies to co-occurring organisms (the 'Black Queen' hypothesis). Genome analyses, moreover, revealed a surprising degree of inter- and intraspecific diversity in metabolic pathways, especially of carbohydrate transport and metabolism, and mainly encoded in genomic islands. The striking genotype microdiversification of acI Actinobacteria might explain their global success in highly dynamic freshwater environments with complex seasonal patterns of allochthonous and autochthonous carbon sources. We propose a new order within Actinobacteria ('Candidatus Nanopelagicales') with two new genera ('Candidatus Nanopelagicus' and 'Candidatus Planktophila') and nine new species.

DOI: https://doi.org/10.1038/ismej.2017.156

Supplementary Information for Neuenschwander et al.

Supplementary text

Figure S1: Microphotographs and cell volumes (μm^3) of isolates. The scale bar at the bottom left applies to all pictures.

Figure S2: Phylogenetic positioning of '*Ca*. Nanopelagicales' based on 16S rRNA genes. a, bootstrapped maximum likelihood tree of 16S rRNA genes; '*Ca.* Aquiluna sp.' and *Rhodoluna lacicola* were used as outgroup. Bootstrap values are shown on the nodes. b, 16S rRNA gene sequence similarity matrix. Species borders are marked with solid lines, genus borders with dashed lines. An asterisk indicates the positioning of the described mixed culture '*Ca*. P. limnetica'.

Figure S3: Average nucleotide identity (ANI) matrix of '*Ca*. Nanopelagicales'. Species borders are marked with solid lines, genus borders with dashed lines.

Figure S4: a, Average amino acid identity (AAI) matrix of '*Ca*. Nanopelagicales'. b, Protein similarity (>50% identity, >50% coverage) matrix of '*Ca*. Nanopelagicales'. Species borders are marked with solid lines, genus borders with dashed lines.

Figure S5: Phylogenomic tree with complete genomes of '*Ca*. Nanopelagicales' only. 462 concatenated conserved proteins were used to generate a maximum-likelihood phylogenetic tree. The genomes of '*Ca.* Aquiluna sp.' and *Rhodoluna lacicola* were used as outgroup. Bootstrap values are indicated by black, grey, and white circles on the nodes, and a colour key is shown on the left.

Figure S6: Phylogenomic tree with complete genomes of the phylum Actinobacteria. Forty-eight concatenated conserved proteins were used to generate a maximum-likelihood phylogenetic tree. The genomes of *Staphylococcus aureus* and *Listeria monocytogenes* were used as outgoup. Bootstrap values are indicated by black, grey or white circles on the nodes, and a colour key is shown on the left. The proposed novel order '*Ca*. Nanopelagicales' is highlighted in green.

Figure S7: Genome streamlining in '*Ca*. Nanopelagicales'. Number of predicted CDS, number of predicted sigma factor homologs, median size of intergenic spacers, and coding density versus genome size for all complete published genomes of Actinobacteria (n=610; data taken from RefSeq). '*Ca*. Nanopelagicales' and *Rhodoluna lacicola* are marked in different colours.

Figure S8: Bootstrapped phylogenetic tree of rhodopsin protein sequences of 'Ca. Nanopelagicales' and other Actinobacteria. Xanthorhodopsin sequences of *Salinibacter ruber* and *Thermus aquaticus* were used as outgroup. Bootstrap values are indicated by coloured circles on the nodes, and a colour key is shown on the left.

Figure S9: Bootstrapped maximum likelihood tree of 23S rRNA genes of '*Ca*. Nanopelagicales'. Target hits for the newly designed probes Pver-23S-1420 and Npel-23S-2669 are shown in brackets. '*Ca.* Aquiluna sp.' and *Rhodoluna lacicola* were used as outgroup. Bootstrap values are shown on the nodes.

Figure S10: Physico-chemical data from Lake Zurich.

Figure S11: Redundancy analysis of environmental parameters explaining the variability in cell numbers of microbes affiliated to all '*Ca*. Nanopelagicales', '*Ca*. Nanopelagicus', and '*Ca*. P. vernalis' in Lake Zurich. temp, water temperature; picocyano, abundance of picocyanobacteria; irrad, irradiation; chloro, chlorophyll *a* associated with chlorophytes; diatom, chlorophyll *a* associated with diatoms; NH4, ammonium concentrations; O2, oxygen concentrations; PO4, phosphate concentrations; NO3, nitrate concentrations; depth, sampling depth.

Figure S12: Metagenomic fragment recruitment of '*Ca*. Nanopelagicus' (a) and '*Ca*. Planktophila' (b) across diverse freshwater ecosystems (see Table S7 for details).

Figure S13: Recruitment plots of time-series metagenomes from Lake Mendota, USA.

Figure S14: Whole-genome alignment of all 13 '*Ca*. Planktophila' strains. The genomes have been linearized for simplicity and are arranged in the same order as in the phylogenomic tree in Fig. 1b. Synteny and different degrees of sequence similarity are indicated by vertical lines connecting the genomes.

Figure S15: Whole-genome alignment of the three '*Ca*. Nanopelagicus' strains. The genomes have been linearized for simplicity and are arranged in the same order as in the phylogenomic tree in Fig. 1b. Synteny and different degrees of sequence similarity are indicated by vertical lines connecting the genomes. rRNA operons in the individual genomes are displayed as red arrows and tRNAs as short vertical lines. Genomic islands (GI) have been marked in different colours and numbered (see Table S10 for genes encoded in each island). Red: genes encoding mainly cell wall biosynthesis and modifications; Yellow: genes encoding mainly membrane transport and / or carbohydrate metabolism.

Table S1: Details of the isolation experiments conducted in Lake Zurich, Switzerland.

Table S2: Preprocessing and assembly parameters.

Table S3: Details of the sequenced strains of planktonic '*Ca*. Nanopelagicales'.

Table S4: Cell size measurements of different strains and *in-situ* in Lake Zurich.

Table S5: Cell sizes and genomic details of genome-streamlined microbes.

Table S6: Details of the applied oligonucleotide probes.

Table S7: Details of the publically available metagenomes that were used for recruitment in Figs. 3, S12, S13. Metagenomes are sorted according to habitat (rivers and lakes) and latitude (separately for North America and Europe). This table is provided as Excel-file (Tables_S7-S10.xlsx).

Table S8: Metabolic pathways in '*Ca.* Nanopelagicales'. This table is provided as Excel-file (Tables_S7-S10.xlsx).

Table S9: Genomic islands in 11 '*Ca*. Planktophila sp.' strains. Locus tags and gene annotations are listed and a general function of the genes encoded in each island is given at the left. This table is provided as Excel-file (Tables_S7-S10.xlsx).

Table S10: Genomic islands in the three '*Ca*. Nanopelagicus sp.' strains. Locus tags and gene annotations are listed and a general function of the genes encoded in each island is given at the left. This table is provided as Excel-file (Tables_S7-S10.xlsx).

Supplementary text

Additional genomic features of '*Ca*. Nanopelagicales'

None of the strains encoded genes for flagella or chemotaxis, confirming the non-motile lifestyle of '*Ca*. Nanopelagicales'. No CRISPR-Cas system was identified, and the number of signal transduction genes was low with only two two-component regulatory systems for phosphate starvation and osmotic stress response and 2-3 additional histidine kinases (Table S7). Nickel superoxide dismutases involved in oxidative stress response were found in all strains, while catalase-peroxidases were present in '*Ca*. Planktophila', but absent in '*Ca*. Nanopelagicus'.

We could confirm the presence of glycolysis via the Embden-Meyerhof pathway, pentose phosphate pathway, tricarboxylic acid (TCA) cycle and oxidative phosphorylation in all genomes, as was proposed from SAGs and MAGs (Garcia *et al.,* 2013, Ghai *et al.,* 2014, Ghylin *et al.,* 2014). Nevertheless, some variations between the different strains were found: '*Ca*. Planktophila' strains encoded a class I fructose-bisphosphate aldolase, while the class II variant was annotated in all '*Ca*. Nanopelagicus' genomes. The former is in contrast to the '*Ca*. Planktophila' genomes sequenced of Kang *et al.* (2017) which all encoded the class II variant. The non-oxidative branch of the pentose phosphate pathway was detected in all genomes, whereas the oxidative branch was only present in '*Ca*. P. dulcis' strains and '*Ca*. P. sulfonica' MMS-IA-56. This is in agreement with previous studies in which genes involved in the oxidative branch of the pentose phosphate pathway were only detected in '*Ca*. Planktophila' (i.e., acI-A / acI-A1) genomes (Ghylin *et al.,* 2014, Kang *et al.,* 2017). Pyruvate dehydrogenase E2 components, involved in the conversion of pyruvate to acetyl-CoA, were only detected in one of the '*Ca*. P. versatilis' strains (MMS-IIB-142). All genomes did however contain genes coding for the 2-oxoglutarate dehydrogenase E2 component, a potential substitute for the latter, which was also the case in the four acI genomes described by Kang *et al.* (2017). Complete gluconeogenesis pathways were detected in '*Ca*. N. limnes' and '*Ca*. N. hibericus' as well as in '*Ca.* P. limnetica' and '*Ca*. P. vernalis', whereas in the other genomes either the first steps (pyruvate to PEP) and/or the last steps (fructose-1,6P_2 to β-D-fructose-6P) were missing. Key enzymes involved in gluconeogenesis have previously been detected in acI-A7, acI-C1 and acI-B1 genomes (Kang *et al.,* 2017, Ghylin *et al.,* 2014) but not in other ac1-A1 genomes. The presence of carbonic anhydrases and PEP carboxylases in all genomes suggests the ability to replenish precursors needed for growth by anapleurotic CO_2 fixation. Previous studies indicated a facultative aerobic lifestyle based on the presence of pathways for pyruvate fermentation (Garcia *et al.,* 2013, Ghai *et al.,* 2014, Ghylin *et al.,* 2014) We cannot confirm this for our genomes. However, all our strains were isolated from 5 m depth from Lake Zurich, where oxic conditions prevail. Oxidative phosphorylation pathways were present in all genomes with the anomaly of missing succinate dehydrogenase subunits SDHC and SDHD as previously described by Kang *et al.*, (2017).

Exopolyphosphatases, high affinity membrane transporters for inorganic phosphate uptake (Pst system), inorganic pyrophosphatases, and two-component regulatory systems for phosphate stress were annotated for all strains, in agreement with predictions from SAGs (Ghylin *et al.,* 2014) and MAGs (Ghai *et al.,* 2014). Phosphorus is usually the limiting element in freshwater systems (Vadstein 2000,

Wetzel 2001), and it seems that '*Ca*. Nanopelagicales have efficient phosphate acquisition systems, although they did not show higher *in-situ* incorporation of inorganic phosphate than other freshwater microbes (Rofner *et al.,* 2016b) and were underrepresented in dissolved organic phosphate uptake (Rofner *et al.,* 2016a). All strains encoded genes for ammonium transport (Amt family), and '*Ca*. P. sulfonica', '*Ca*. P. versatilis' and '*Ca*. P. lacus' strains had two copies. Transporters for other N-rich components like polyamines (i.e., spermidine/putrescine) and amino acids as well as cyanophycinases were predicted for all strains and cyanate transporters were present in '*Ca*. P. versatilis' and '*Ca*. P. vernalis'. This high prevalence of genes involved in the uptake of organic N-rich compounds appears to be a key characteristic for '*Ca*. Nanopelagicales' (Garcia *et al.,* 2013*;* Ghai *et al.,* 2014; Kang *et al.,* 2017). A high *in-situ* uptake of different amino acids was previously reported from MAR-FISH assays, as well as high glucose incorporation (Buck *et al.,* 2009, Pérez *et al.,* 2015, Salcher *et al.,* 2010, Salcher *et al.,* 2013). Polyamine degradation pathways were present in all genomes, however, not complete. With the exception of '*Ca.* P. sulfonica' all strains were able to convert 4-aminobutyraldehyde to succinate; while enzymes catalysing the first three steps of the spermidine/putrescine degradation pathway were not annotated. Most genomes, however, contained acetylornithine aminotransferases which belong to the same superfamily as putrescine aminotransferases, and thus, might fulfil this function. Lysozymes were predicted in all genomes, as well as putative chitinases (GH18 family, glycoside hydrolase CAZy; (Garcia *et al.,* 2013)) that include signal sequences for membrane transport, consistent with SAGs (Ghylin *et al.,* 2014). Their low homology to chitinases originating from other than acI SAGs or MAGs, however, makes interpretation difficult, particularly since genes involved in amino-sugar degradation were only detected in '*Ca*. P. dulcis' and '*Ca*. P. sulfonica'. We could also not identify transporters for amino sugars although '*Ca*. Nanopelagicales' are known to take up N-acetylglucosamine *in situ* (Beier and Bertilsson 2011, Eckert *et al.,* 2012). Other transporters, especially for carbohydrates, as well as genes encoding carbohydrate breakdown, were present in a highly variable fashion in the genomes: '*Ca*. P. dulcis' and '*Ca*. P. vernalis' strains had the highest number of different carbohydrate transporters and enzymes involved in carbohydrate metabolism, while '*Ca*. Nanopelagicus' strains and '*Ca*. P. limnetica' were less versatile in carbohydrate usage (Tables 2, S7). Membrane transporters for sulfonate and benzoate were annotated in two strains only ('*Ca*. P. sulfonica' and 'Ca. P. limnetica', respectively). Transporters for biotin (*bioY*) and energy-coupling factor transporters with preceding AdoCbl riboswitches for cobalamin were annotated in all '*Ca*. Planktophila' strains except for '*Ca*. P. limnetica'. Although transporters for thiamine and riboflavin were not found, additional NAD transporters (*pnuC*) preceded by riboswitches for thiamine and/or riboflavin were annotated in all '*Ca*. Nanopelagicus'. This is in accordance with other genomes associated with '*Ca*. Nanopelagicales' (Kang *et al.,* 2017). As Pnu transporters share high homologies to each other, it is likely that the putative PnuC either function as thiamine (PnuT) or riboflavin (PnuX) transporters (Jaehme and Slotboom Dirk 2015).

Description of the proposed '*Candidatus'* taxa

Based on our analysis we suggest that our strains represent a novel '*Candidatus'* order and '*Candidatus'* family in the phylum Actinobacteria with two new genera and nine new species.

‘*Candidatus* Nanopelagicales’ [Na.no.pe.la.gi.ca’les. N.L. masc. n. *Nanopelagicus* type genus of the order; suff. -*ales*, ending to denote an order; N.L. fem. pl. n. *Nanopelagicales*, the order of the genus *Nanopelagicus*]. Aerobic chemoheterotrophs. Cells are tiny, non-motile, and inhabit the plankton of freshwaters. ‘*Ca*. Nanopelagicales’ can be recognized by the presence of the diagnostic oligonucleotide sequence 5’-AATGCGTTAGCTGCGTCGCA-3’ in the 16S rRNA gene (positions 852-872, *E. coli* numbering). A member of the phylum Actinobacteria. Contains a single family, ‘*Candidatus* Nanopelagicus’, and two genera ‘*Candidatus* Nanopelagicus’ and ‘*Candidatus* Planktophila’. Basis of the assignment is a phylogenetic tree of 48 conserved concatenated proteins of >100 complete genomes of all orders of Actinobacteria (Fig. 2, S6).

‘*Candidatus* Nanopelagicaceae’ [Na.no.pe.la.gi.ca.ce’ae. N.L. masc. n. *Nanopelagicus* type genus of the family; suff. -*aceae*, ending to denote a family; N.L. fem. pl. n. *Nanopelagicaceae*, the order of genus *Nanopelagicus*.]

‘*Candidatus* Nanopelagicus limnes’ [Na.no.pe.la.gi.cus. N.L. masc. n. *nano* very small; L. masc. adj. *pelagicus* belonging to the pelagial; N.L. masc. n. *Nanopelagicus* very small pelagic; referring to the small cell and genome size and the pelagic habitat; lim’nes. L. gen. n. *limnes* of a lake]. Represented by strain MMS-21-122, which was isolated from Lake Zurich, Switzerland. Curved rods with lengths of 0.45±0.09 µm and diameters of 0.25±0.03 µm (Table S3, Figure S1). The initial pure culture was lost after a few propagations to fresh medium; no growing culture is available. The initial culture grew well in sterile lake water amended with minimal carbon medium, vitamins and amino acids. ‘*Ca*. N. limnes’ MMS-21-122 has a genome size of 1.24 Mbp and a genomic GC content of 41.5 %. It is auxotrophic for reduced sulfur sources, several amino acids (proline, ornithine, histidine, betaine) and several vitamins (B1, B2, B5, B7, B12) and possesses rhodopsins (Table S7). Members of the genus ‘*Ca*. N. abundans’ can be recognized by the presence of the diagnostic oligonucleotide sequence 5’-ACAAGAGGTTCGTCCGTCC-3’ in the 23S rRNA gene (positions 2669-2688, *E. coli* numbering). A complete genome of ‘*Ca*. N. limnes’ MMS-21-122 is available at Genbank (CP016768). Phylogenetic analyses of 476 concatenated protein sequences, 16S rRNA, 23S rRNA, and rhodpdopsin genes, as well as average nucleotide identities and protein similarities indicated that the *Candidatus* taxon belongs to a novel genus (‘*Ca.* Nanopelagicus’) of the family ‘*Ca*. Nanopelagicaceae’ and the novel order ‘*Ca.* Nanopelagicales’.

‘*Candidatus* Nanopelagicus hibericus’ [hi.be’ri.cus. L. masc. adj. *hibericus* Spanish; referring to a high abundance in two Spanish reservoirs]. Represented by strain MMS-21-160, which was isolated from Lake Zurich, Switzerland. Rods with lengths of 0.33±0.07 µm and diameters of 0.24±0.03 µm (Table S3, Figure S1). The initial pure culture was lost after a few propagations to fresh medium; no growing culture is available. The initial culture grew well in sterile lake water amended with minimal carbon medium, vitamins and amino acids. ‘*Ca*. N. hibericus’ MMS-21-160 has a genome size of 1.22 Mbp and a genomic GC content of 42.4 %. It is auxotrophic for reduced sulfur sources, several amino acids (ornithine, histidine, betaine) and several vitamins (B1, B2, B5, B7, B12) and possesses rhodopsins (Table S7). A complete genome is available at Genbank (CP016771). Phylogenetic analyses of 476 concatenated protein sequences, 16S rRNA, 23S rRNA, and rhodpdopsin genes, as well as average nucleotide identities and protein similarities indicated the affiliation of the *Candidatus* taxon to the novel genus ‘*Ca.*

Nanopelagicus' of the family '*Ca*. Nanopelagicaceae' and the novel order '*Ca.* Nanopelagicales'.

'*Candidatus* Nanopelagicus abundans' [a.bun'dans. L. pres. part. *abundans* abundant; referring to high global abundances]. Represented by strain MMS-IIB-91, which was isolated from Lake Zurich, Switzerland. Rods with lengths of 0.46±0.47 µm and diameters of 0.26±0.20 µm (Table S3, Figure S1). The initial pure culture was lost after a few propagations to fresh medium; no growing culture is available. The initial culture grew well in sterile lake water amended with inorganic basal medium, minimal carbon medium, vitamins and amino acids. '*Ca*. N. abundans' MMS-IIB-91 has a genome size of 1.16 Mbp and a genomic GC content of 40.2 %. It is auxotrophic for reduced sulfur sources, several amino acids (methionine, lysine, ornithine, histidine, betaine), several vitamins (B1, B5, B7, B12), and possesses rhodopsins (Table S7). A complete genome is available at Genbank (CP016779). Phylogenetic analyses of 476 concatenated protein sequences, 16S rRNA, 23S rRNA, and rhodpdopsin genes, as well as average nucleotide identities and protein similarities indicated the affiliation of the *Candidatus* taxon to the novel genus '*Ca.* Nanopelagicus' of the family '*Ca*. Nanopelagicaceae' and the novel order '*Ca.* Nanopelagicales'.

'*Candidatus* Planktophila limnetica' [Plank.to'phi.la. N.L.neut. n. *plankton* plankton; N.L. fem. adj. *phila* friendly to;N.L. fem. n. *Planktophila* the friend (fem.) of plankton; lim.ne'ti.ca. N.L. fem. adj. *limnetica* pertaining to lakes]. According to the previous description of a mixed culture (MWH-EgelM2-3.acI)(Jezbera *et al.,* 2009) that has a 100% identical 16S rRNA gene sequences with strain MMS-VB-114, we propose the name '*Ca*. P. limnetica' for strain MMS-VB-114 and make an amended description based on a complete genome. Represented by strain MMS-VB-114, which was isolated from Lake Zurich, Switzerland. Curved rods with lengths of 0.38±0.10 µm and diameters of 0.25±0.06 µm (Table S3, Figure S1). The initial pure culture was lost after a few propagations to fresh medium; no growing culture is available. The initial culture grew well in sterile lake water amended with pyruvate, urea, vitamins, and amino acids. '*Ca*. P. limnetica' MMS-VB-114 has a genome size of 1.33 Mbp and a genomic GC content of 45.0 %. It is auxotrophic for reduced sulfur sources and vitamins (B1, B5, B7), and possesses rhodopsins (Table S7). A complete genome is available at Genbank (CP016782). Phylogenetic analyses of 476 concatenated protein sequences, 16S rRNA, 23S rRNA, and rhodopsin genes, as well as average nucleotide identities and protein similarities indicated the affiliation of the *Candidatus* taxon to the family '*Ca*. Nanopelagicaceae' and the novel order '*Ca.* Nanopelagicales'.

'*Candidatus* Planktophila dulcis' [dul'cis. L. fem. adj. *dulcis* sweet; referring to a high diversity of sugar transporters and metabolism]. Represented by strain MMS-IIA-65, which was isolated from Lake Zurich, Switzerland. Curved rods with lengths of 0.44±0.09 µm and diameters of 0.26±0.03 µm (Table S3, Figure S1). The initial pure culture was lost after a few propagations to fresh medium; no growing culture is available. The initial culture grew well in sterile lake water amended with pyruvate, urea, minimal carbon medium, vitamins, and amino acids. '*Ca*. P. dulcis' MMS-IIA-65 has a genome size of 1.35 Mbp and a genomic GC content of 48.0 %. It is auxotrophic for reduced sulfur sources, betaine, and several vitamins (B1, B5, B7, B12), and possesses rhodopsins (Table S7). It has a high number of different carbohydrate transporters. A complete genome is available at Genbank (CP016777).

Phylogenetic analyses of 476 concatenated protein sequences, 16S rRNA, 23S rRNA, and rhodopsin genes, as well as average nucleotide identities and protein similarities indicated the affiliation of the *Candidatus* taxon to the genus '*Ca.* Planktophila' of the family '*Ca*. Nanopelagicaceae' and the novel order '*Ca.* Nanopelagicales'. Two more strains (MMS-IA-53 and MMS-21-155) isolated from Lake Zurich share high similarities with strain MMS-IIA-65 and are thus affiliated to the same *Candidatus* species. This assignment is based on similar cell and genome sizes, similar metabolism, phylogenetic analyses of 476 concatenated protein sequences, 16S rRNA (100% identical), 23S rRNA, and rhodopsin genes, as well as average nucleotide identities (<97%) and protein similarities (<94%). Complete genomes of '*Ca*. P. dulcis' MMS-IA-53 and MMS-21-155 are available at Genbank (CP016772, CP016770).

'*Candidatus* Planktophila sulfonica' [sul.fo'ni.ca. N.L. fem. adj. *sulfonica* pertaining to sulfonate; referring to sulfonate transporters]. Represented by strain MMS-IA-56, which was isolated from Lake Zurich, Switzerland. Curved rods with lengths of 0.50±0.14 µm and diameters of 0.28±0.06 µm (Table S3, Figure S1). The initial pure culture was lost after a few propagations to fresh medium; no growing culture is available. The initial culture grew well in sterile lake water amended with inorganic basal medium, minimal carbon medium, vitamins, and amino acids. It is so far the only member of the order '*Ca*. Nanopelagicales' that possesses membrane transporters for sulfonates. '*Ca*. P. sulfonica' MMS-IA-56 has a genome size of 1.34 Mbp and a genomic GC content of 48.6 %. It is auxotrophic for reduced sulfur sources, betaine, and several vitamins (B1, B5, B7, B12), and possesses rhodopsins (Table S7). A complete genome is available at Genbank (CP016773). Phylogenetic analyses of 476 concatenated protein sequences, 16S rRNA, 23S rRNA, and rhodopsin genes, as well as average nucleotide identities and protein similarities indicated the affiliation of the *Candidatus* taxon to the genus '*Ca.* Planktophila' of the family '*Ca*. Nanopelagicaceae' and the novel order '*Ca.* Nanopelagicales'.

'*Candidatus* Planktophila versatilis' [ver.sa'ti.lis. L. fem. adj. *versatilis* versatile; referring to high metabolic versatility]. Represented by strain MMS-IA-79, which was isolated from Lake Zurich, Switzerland. Curved rods with lengths of 0.45±0.10 µm and diameters of 0.27±0.04 µm (Table S3, Figure S1). The initial pure culture was lost after a few propagations to fresh medium; no growing culture is available. The initial culture grew well in sterile lake water amended with inorganic basal medium, minimal carbon medium, vitamins, and amino acids. '*Ca*. P. versatilis' MMS-IA-79 has a genome size of 1.33 Mbp and a genomic GC content of 48.2 %. It is auxotrophic for reduced sulfur sources, betaine, and several vitamins (B1, B5, B7, B12), and possesses rhodopsins (Table S7). A complete genome is available at Genbank (CP016778). Phylogenetic analyses of 476 concatenated protein sequences, 16S rRNA, 23S rRNA, and rhodopsin genes, as well as average nucleotide identities and protein similarities indicated the affiliation of the *Candidatus* taxon to the genus '*Ca.* Planktophila' of the family '*Ca*. Nanopelagicaceae' and the novel order '*Ca.* Nanopelagicales'. Three additional strains (MMS-IIB-79, MMS-IA-105, and MMS-IIB-142) isolated from Lake Zurich share high similarities with strain MMS-IA-79 and are thus affiliated to the same *Candidatus* species. This assignment is based on similar cell and genome sizes, similar metabolism, phylogenetic analyses of 476 concatenated protein sequences, 16S rRNA (100% identical), 23S rRNA, and rhodopsin genes, as well as average nucleotide identities (<95%) and protein similarities (<90%). Complete genomes of '*Ca*. P. versatilis' MMS-IIB-79, MMS-IA-

105, and MMS-IIB-142 are available at Genbank (CP016774, CP016775, CP016781).

'*Candidatus* Planktophila lacus' [la'cus. L. gen. masc. n. *lacus* of a lake]. Represented by strain MMS-21-148, which was isolated from Lake Zurich, Switzerland. Rods with lengths of 0.41±0.10 µm and diameters of 0.30±0.06 µm (Table S3, Figure S1). The initial pure culture was lost after a few propagations to fresh medium; no growing culture is available. The initial culture grew well in sterile lake water amended with minimal carbon medium, vitamins, and amino acids. '*Ca*. P. lacus' MMS-21-148 has a genome size of 1.46 Mbp and a genomic GC content of 47.5 %. It is auxotrophic for reduced sulfur sources, betaine, and several vitamins (B1, B5, B7, B12), and possesses rhodopsins (Table S7). A complete genome is available at Genbank (CP016769). Phylogenetic analyses of 476 concatenated protein sequences, 16S rRNA, 23S rRNA, and rhodopsin genes, as well as average nucleotide identities and protein similarities indicated the affiliation of the *Candidatus* taxon to the genus '*Ca.* Planktophila' of the family '*Ca*. Nanopelagicaceae' and the novel order '*Ca.* Nanopelagicales'. Two additional strains (MMS-IIB-106 and MMS-IIB-60) isolated from Lake Zurich share high similarities with strain MMS-21-148 and are thus affiliated to the same *Candidatus* species. This assignment is based on similar cell and genome sizes, similar metabolism, phylogenetic analyses of 476 concatenated protein sequences, 16S rRNA (100% identical), 23S rRNA, and rhodopsin genes, as well as average nucleotide identities (<94%) and protein similarities (<95%). Complete genomes of '*Ca*. P. lacus' MMS-IIB-60 and MMS-IIB-106 are available at Genbank (CP016780, CP016783).

'*Candidatus* Planktophila vernalis' [ver.na'lis. L. fem. adj. *vernalis* belonging to spring; referring to high abundances in spring]. Represented by strain MMS-IIA-15, which was isolated from Lake Zurich, Switzerland. Curved rods with lengths of 0.43±0.12 µm and diameters of 0.26±0.05 µm (Table S3, Figure S1). The initial pure culture was lost after a few propagations to fresh medium; no growing culture is available. The initial culture grew well in sterile lake water amended with pyruvate, urea, minimal carbon medium, vitamins, and amino acids. '*Ca*. P. vernalis' MMS-IIA-15 has a genome size of 1.36 Mbp and a genomic GC content of 45.7 %. It is auxotrophic for reduced sulfur sources, serine, and several vitamins (B1, B5, B7, B12), and possesses rhodopsins (Table S7). '*Ca*. P. vernalis' MMS-IIA-15 can be recognized by the presence of the diagnostic oligonucleotide sequence 5'-AACTACTACCACACCGGTTCG-3' in the 23S rRNA gene (positions 1420-1441, *E. coli* numbering). A complete genome is available at Genbank (CP016776). Phylogenetic analyses of 476 concatenated protein sequences, 16S rRNA, 23S rRNA, and rhodopsin genes, as well as average nucleotide identities and protein similarities indicated the affiliation of the *Candidatus* taxon to the genus '*Ca.* Planktophila' of the family '*Ca*. Nanopelagicaceae' and the novel order '*Ca.* Nanopelagicales'.

References:

Beier S, Bertilsson S (2011). Uncoupling of chitinase activity and uptake of hydrolysis products in freshwater bacterioplankton. *Limnol Oceanogr* **56:** 1179-1188.

Buck U, Grossart HP, Amann R, Pernthaler J (2009). Substrate incorporation patterns of bacterioplankton populations in stratified and mixed waters of a humic lake. *Environ Microbiol* **11:** 1854-1865.

Eckert EM, Salcher MM, Posch T, Eugster B, Pernthaler J (2012). Rapid successions affect microbial N-acetyl-glucosamine uptake patterns during a lacustrine spring phytoplankton bloom. *Environ Microbiol* **14:** 794-806.

Garcia SL, McMahon KD, Martinez-Garcia M, Srivastava A, Sczyrba A, Stepanauskas R *et al* (2013). Metabolic potential of a single cell belonging to one of the most abundant lineages in freshwater bacterioplankton. *ISME J* **7:** 137-147.

Ghai R, Mizuno CM, Picazo A, Camacho A, Rodriguez-Valera F (2014). Key roles for freshwater Actinobacteria revealed by deep metagenomic sequencing. *Molecular Ecology* **23:** 6073-6090.

Ghylin TW, Garcia SL, Moya F, Oyserman BO, Schwientek P, Forest KT *et al* (2014). Comparative single-cell genomics reveals potential ecological niches for the freshwater acI Actinobacteria lineage. *ISME J* **8:** 2503-2516.

Jaehme M, Slotboom Dirk J (2015). Structure, function, evolution, and application of bacterial Pnu-type vitamin transporters. *Biological Chemistry*. p 955.

Jezbera J, Sharma AK, Brandt U, Doolittle WF, Hahn MW (2009). '*Candidatus* Planktophila limnetica', an actinobacterium representing one of the most numerically important taxa in freshwater bacterioplankton. *Int J Syst Evol Microbiol* **59:** 2864-2869.

Kang I, Ki m S, Islam MR, Cho J-C (2017). The first complete genome sequences of the acI lineage, the most abundant freshwater Actinobacteria, obtained by whole-genome-amplification of dilution-to-extinction cultures. *Scientific Reports* **7:** 42252.

Pérez MT, Rofner C, Sommaruga R (2015). Dissolved organic monomer partitioning among bacterial groups in two oligotrophic lakes. *Environ Microbiol Rep* **7:** 265–272.

Rofner C, Sommaruga R, Pérez MT (2016a). Differential utilization patterns of dissolved organic phosphorus compounds by heterotrophic bacteria in two mountain lakes. *FEMS Microbiol Ecol* **92:** fiw139.

Rofner C, Sommaruga R, Teresa Pérez M (2016b). Phosphate and ATP uptake by lake bacteria: does taxonomical identity matter? *Environ Microbiol* **18:** 4782-4793.

Salcher MM, Pernthaler J, Posch T (2010). Spatiotemporal distribution and activity patterns of bacteria from three phylogenetic groups in an oligomesotrophic lake. *Limnol Oceanogr* **55:** 846-856.

Salcher MM, Posch T, Pernthaler J (2013). *In situ* substrate preferences of abundant bacterioplankton populations in a prealpine freshwater lake. *ISME J* **7:** 896-907.

Vadstein O (2000). Heterotrophic, planktonic bacteria and cycling phosphorus. Phosphorus requirements, competitive ability, and food web interactions. *Advances in Microb Ecol* **16:** 115-167.

Wetzel R (2001). *Limnology. Lake and River Ecosystems*, 3rd edn. Elsevier Academic Press.

Figure S1: Microphotographs and cell volumes (µm^3) of isolates. The scale bar at the bottom left applies to all pictures.

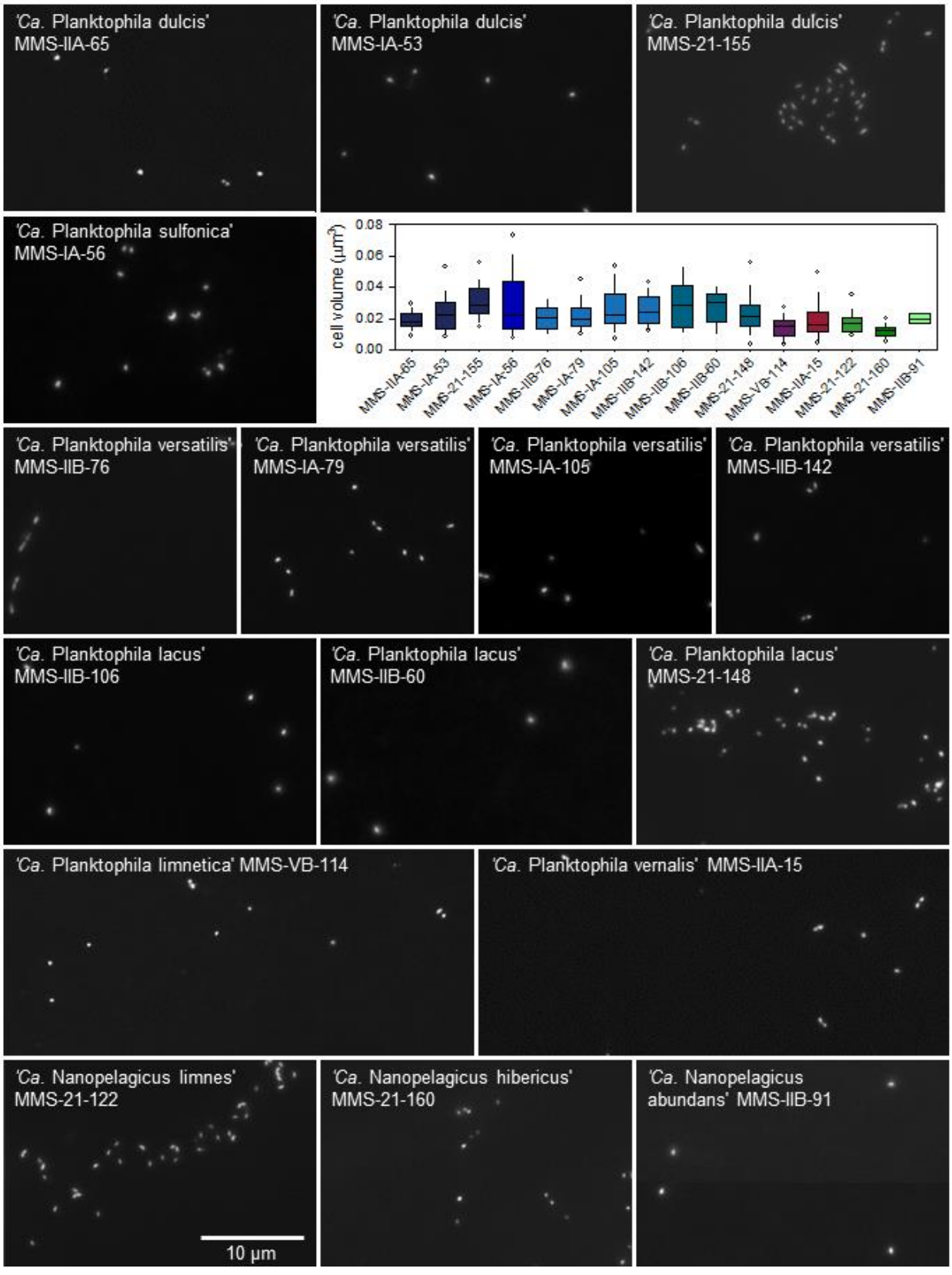


Figure S2: Phylogenetic positioning of '*Ca*. Nanopelagicales' based on 16S rRNA genes. a, bootstrapped maximum likelihood tree of 16S rRNA genes; '*Ca.* Aquiluna sp.' and *Rhodoluna lacicola* were used as outgroup. Bootstrap values are shown on the nodes. b, 16S rRNA gene sequence similarity matrix. Species borders are marked with solid lines, genus borders with dashed lines. An asterisk indicates the positioning of the described mixed culture '*Ca*. P. limnetica'.

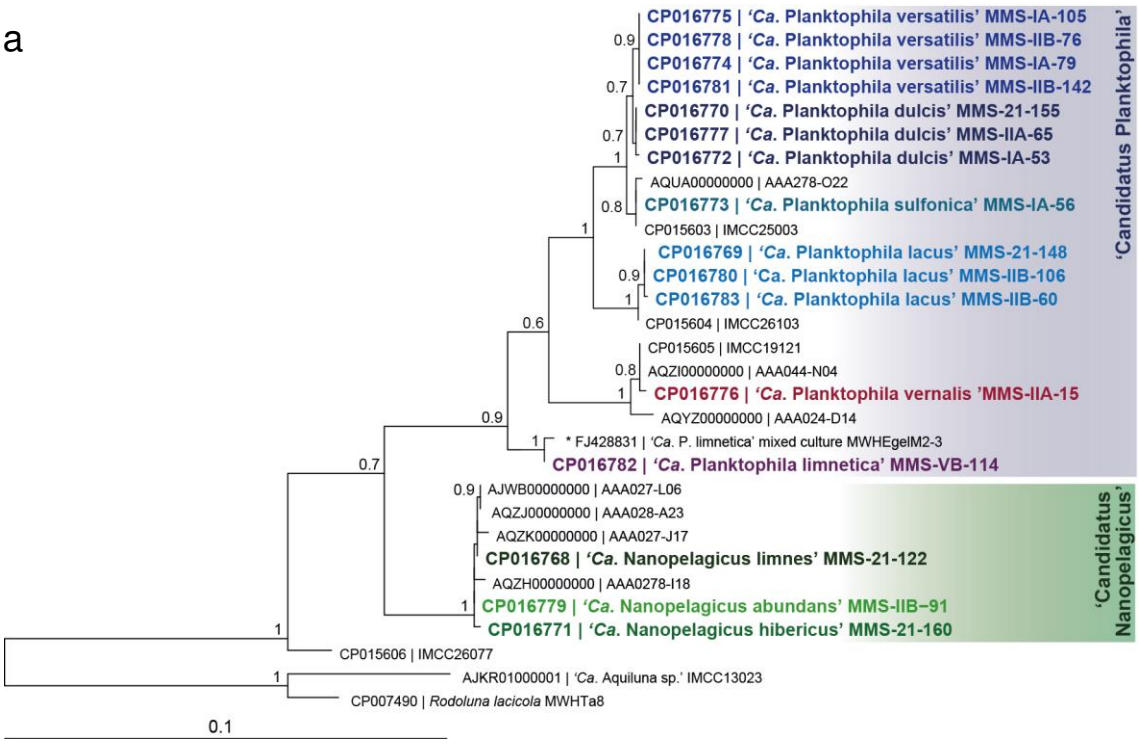


Figure S3: Average nucleotide identity (ANI) matrix of '*Ca*. Nanopelagicales'. Species borders are marked with solid lines, genus borders with dashed lines.

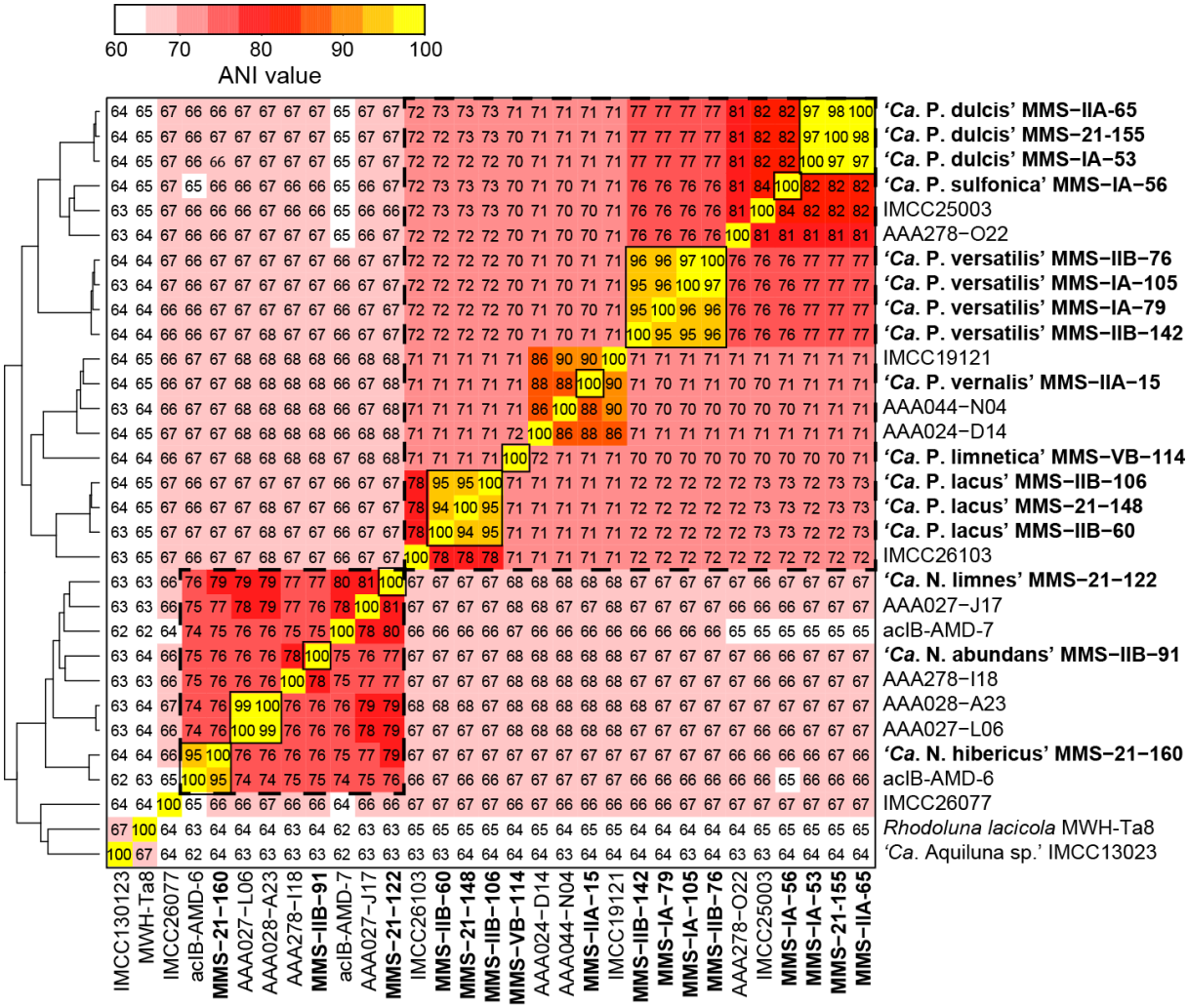
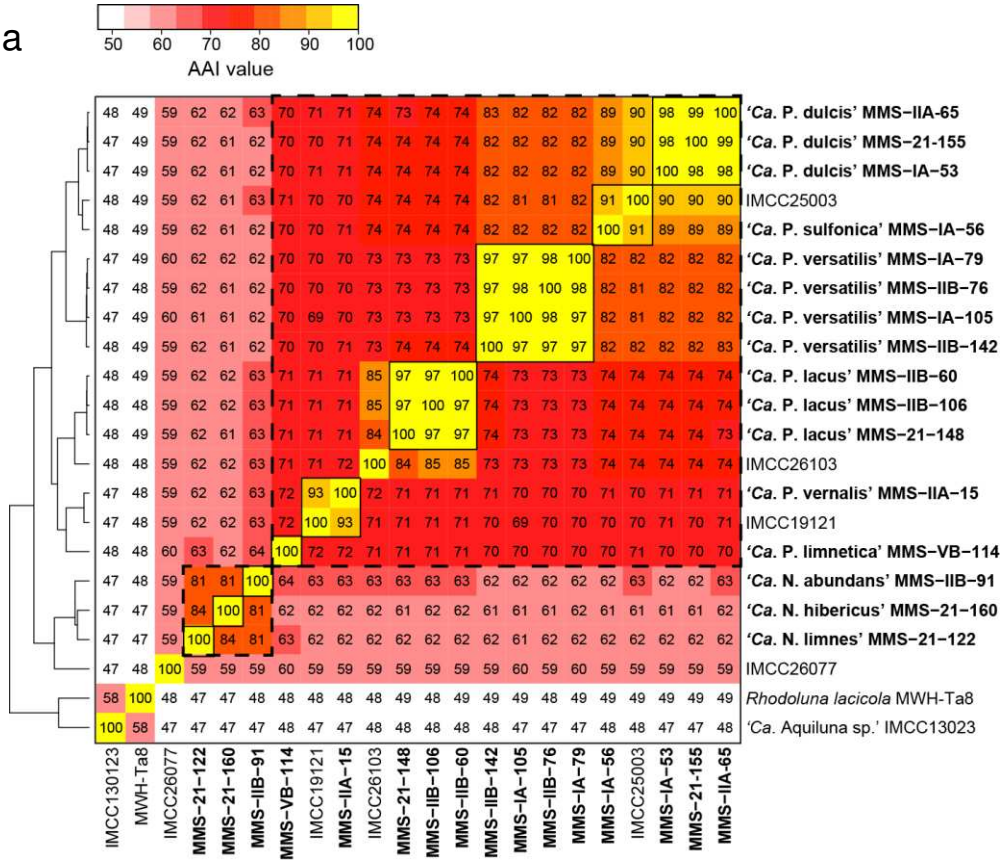


Figure S4: a, Average amino acid identity (AAI) matrix of '*Ca.* Nanopelagicales'. b, Protein similarity (>50% identity, >50% coverage) matrix of '*Ca.* Nanopelagicales'. Species borders are marked with solid lines, genus borders with dashed lines.

a



b

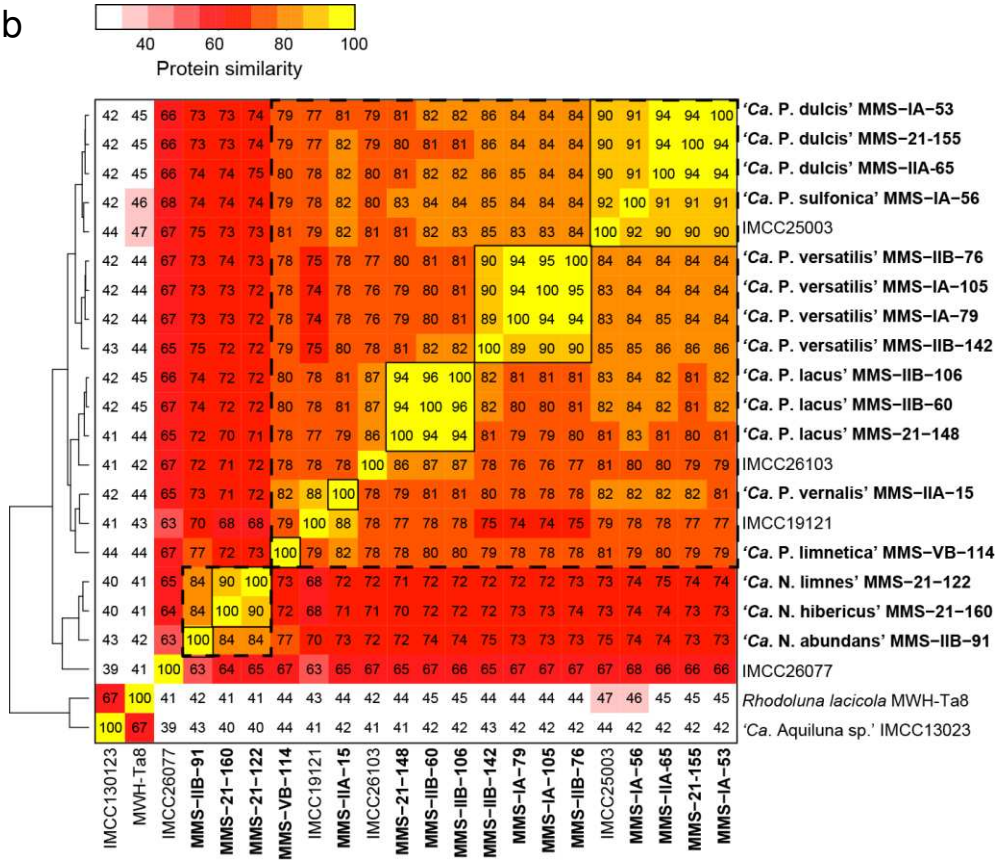


Figure S5: Phylogenomic tree with complete genomes of '*Ca.* Nanopelagicales' only. 462 concatenated conserved proteins were used to generate a maximum-likelihood phylogenetic tree. The genomes of '*Ca.* Aquiluna sp.' and *Rhodoluna lacicola* were used as outgroup. Bootstrap values are indicated by black, grey, and white circles on the nodes, and a colour key is shown on the left.

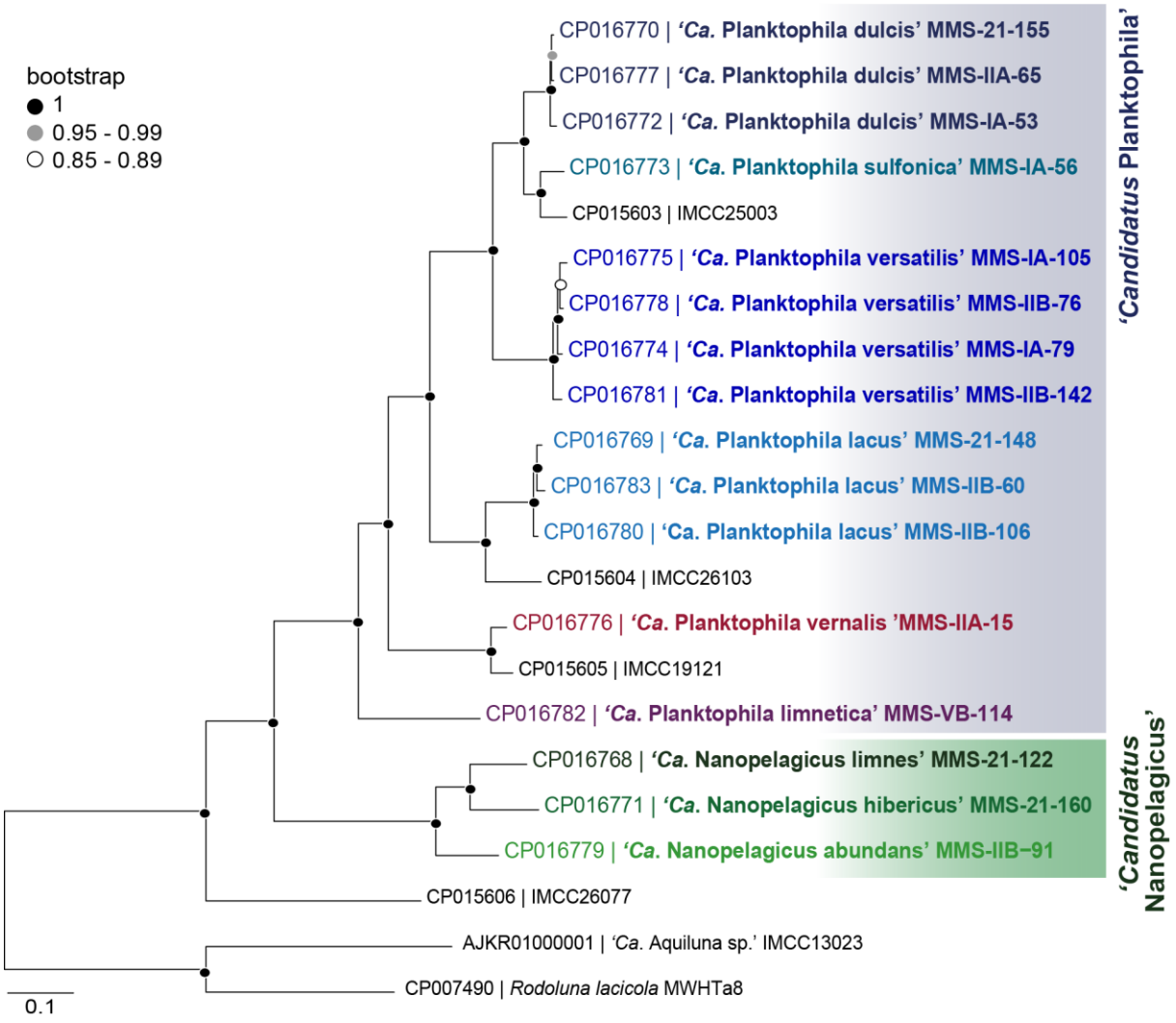


Figure S6: Phylogenomic tree with complete genomes of the phylum Actinobacteria. Forty-eight concatenated conserved proteins were used to generate a maximum-likelihood phylogenetic tree. The genomes of *Staphylococcus aureus* and *Listeria monocytogenes* were used as outgoup. Bootstrap values are indicated by black, grey or white circles on the nodes, and a colour key is shown on the left. The proposed novel order '*Ca*. Nanopelagicales' is highlighted in green.

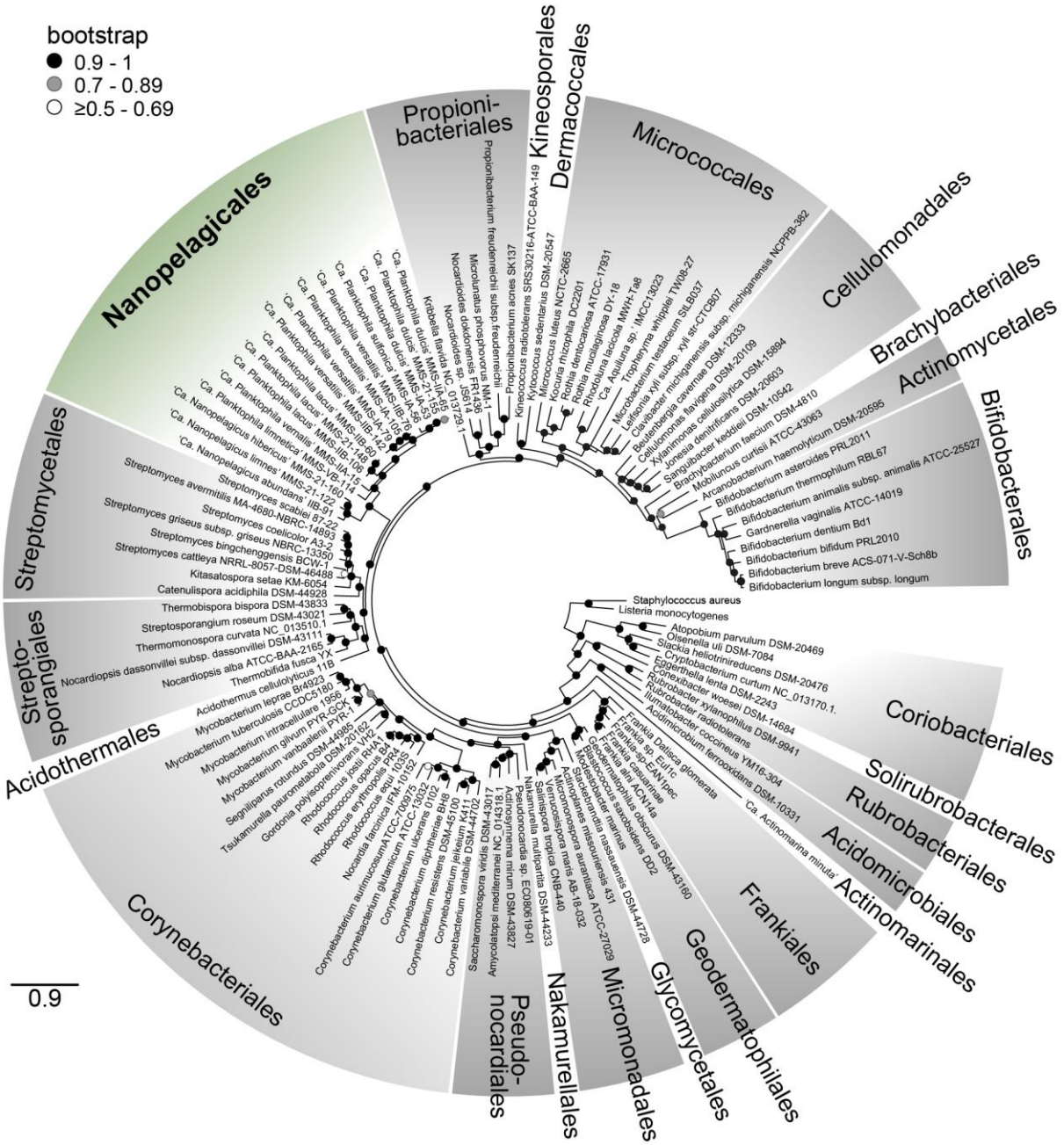


Figure S7: Genome streamlining in '*Ca*. Nanopelagicales'. Number of predicted CDS, number of predicted sigma factor homologs, median size of intergenic spacers, and coding density versus genome size for all complete published genomes of Actinobacteria (n=610; data taken from RefSeq). '*Ca*. Nanopelagicales' and *Rhodoluna lacicola* are marked in different colours.

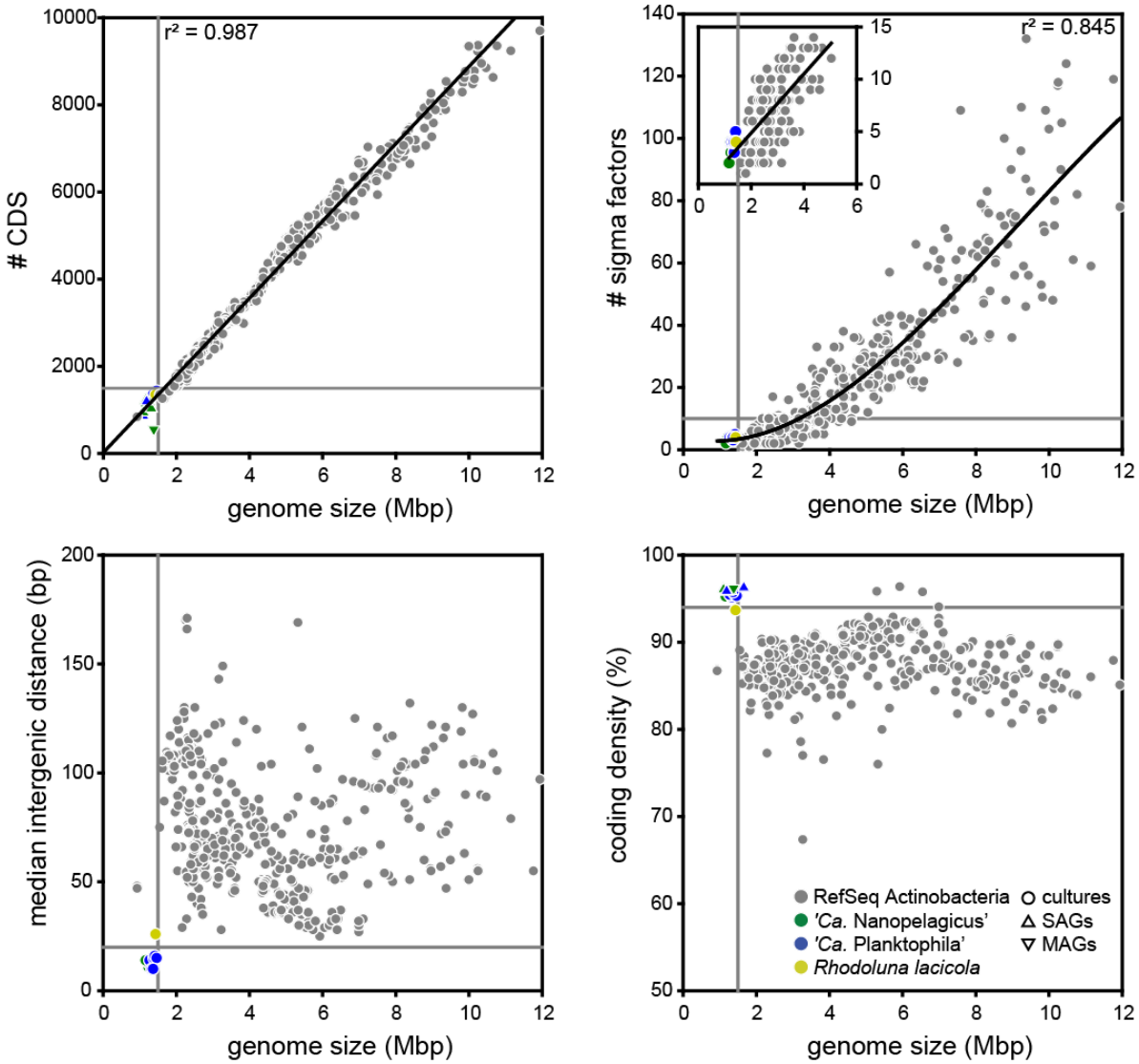


Figure S8: a, Bootstrapped phylogenetic tree of rhodopsin protein sequences of '*Ca.* Nanopelagicales' and other Actinobacteria. Xanthorhodopsin sequences of *Salinibacter ruber* and *Thermus aquaticus* were used as outgroup. Bootstrap values are indicated by coloured circles on the nodes, and a colour key is shown on the left. b, Arrangement of actinorhodopsin (AR) and carotenoid synthesis genes, i.e., beta carotene (*crtE, crtI, crtB, crtY*) and retinal (*blh*) biosynthesis genes. Genomes are arranged according to the phylogenomic tree in Fig. S5.

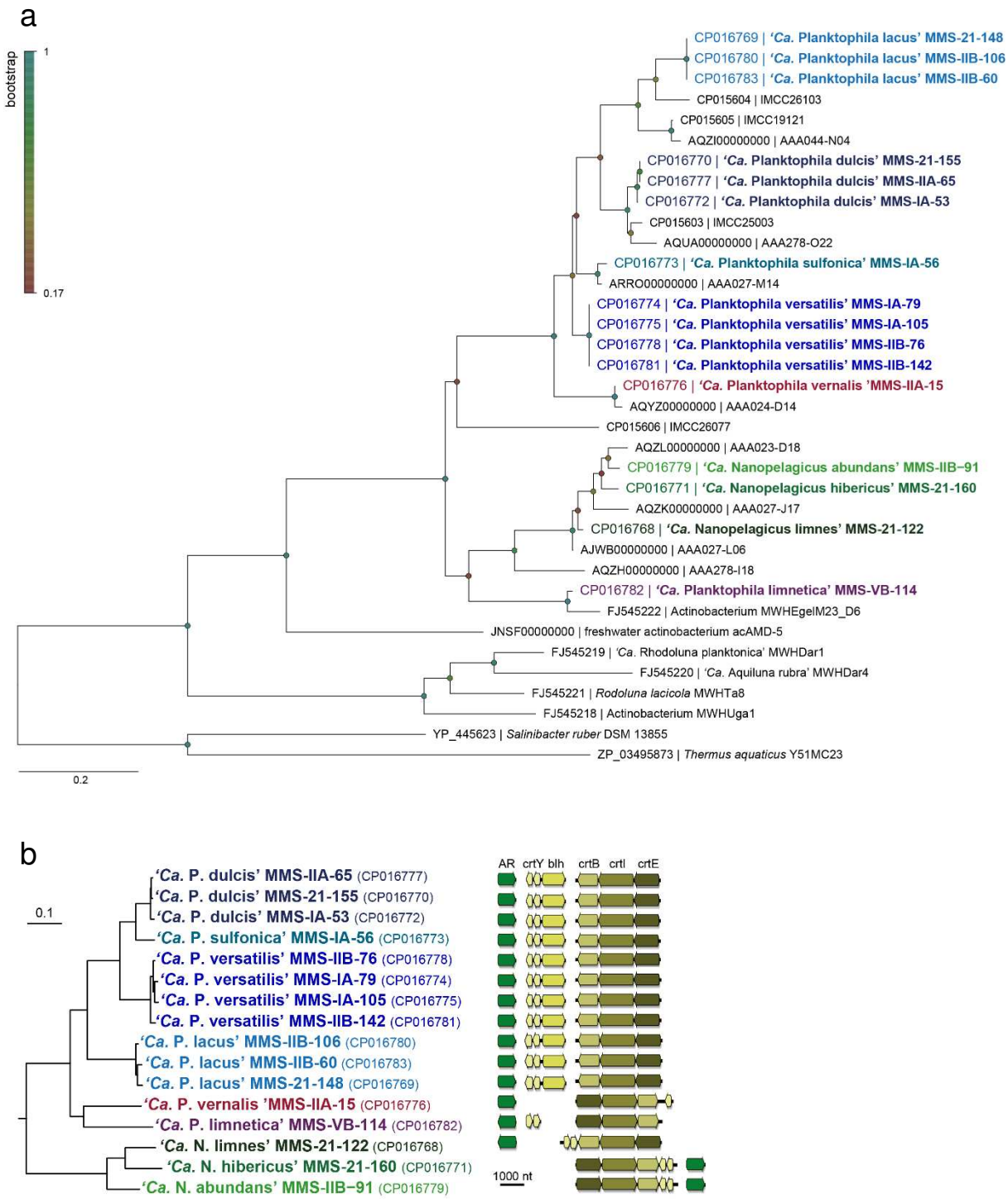


Figure S9: Bootstrapped maximum likelihood tree of 23S rRNA genes of '*Ca.* Nanopelagicales'. Target hits for the newly designed probes Pver-23S-1420 and Npel-23S-2669 are shown in brackets. '*Ca.* Aquiluna sp.' and *Rhodoluna lacicola* were used as outgroup. Bootstrap values are shown on the nodes.

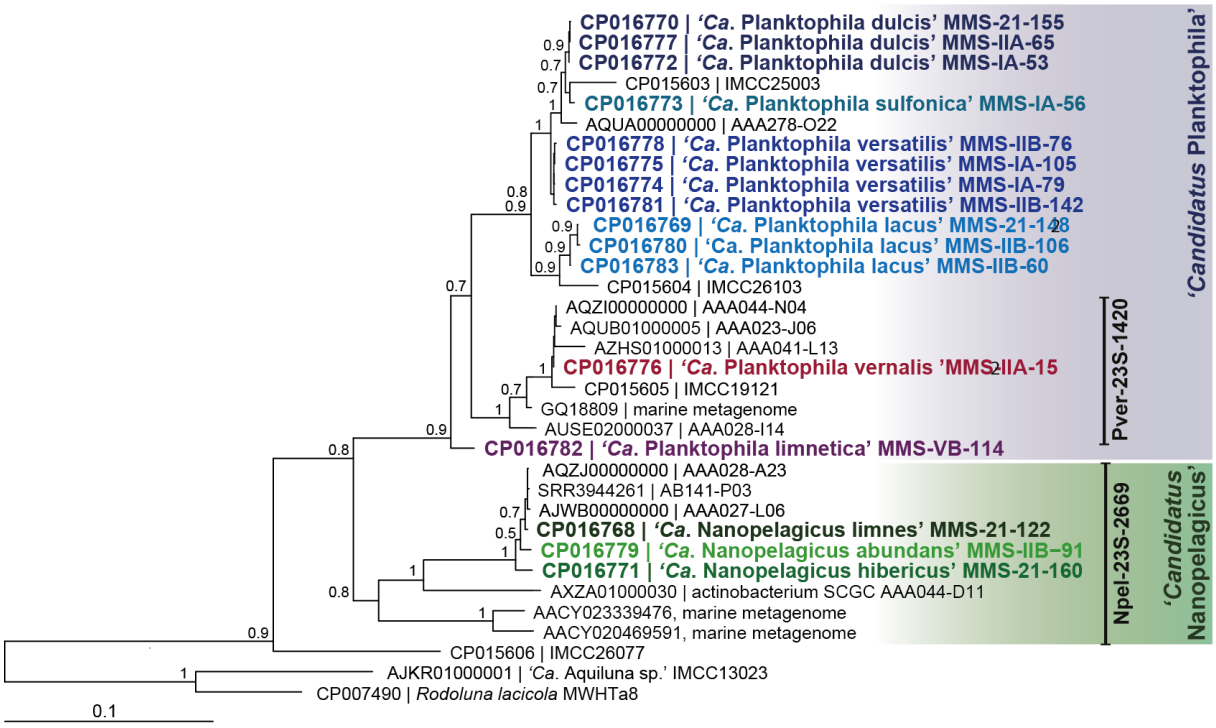


Figure S10: Physico-chemical data from Lake Zurich

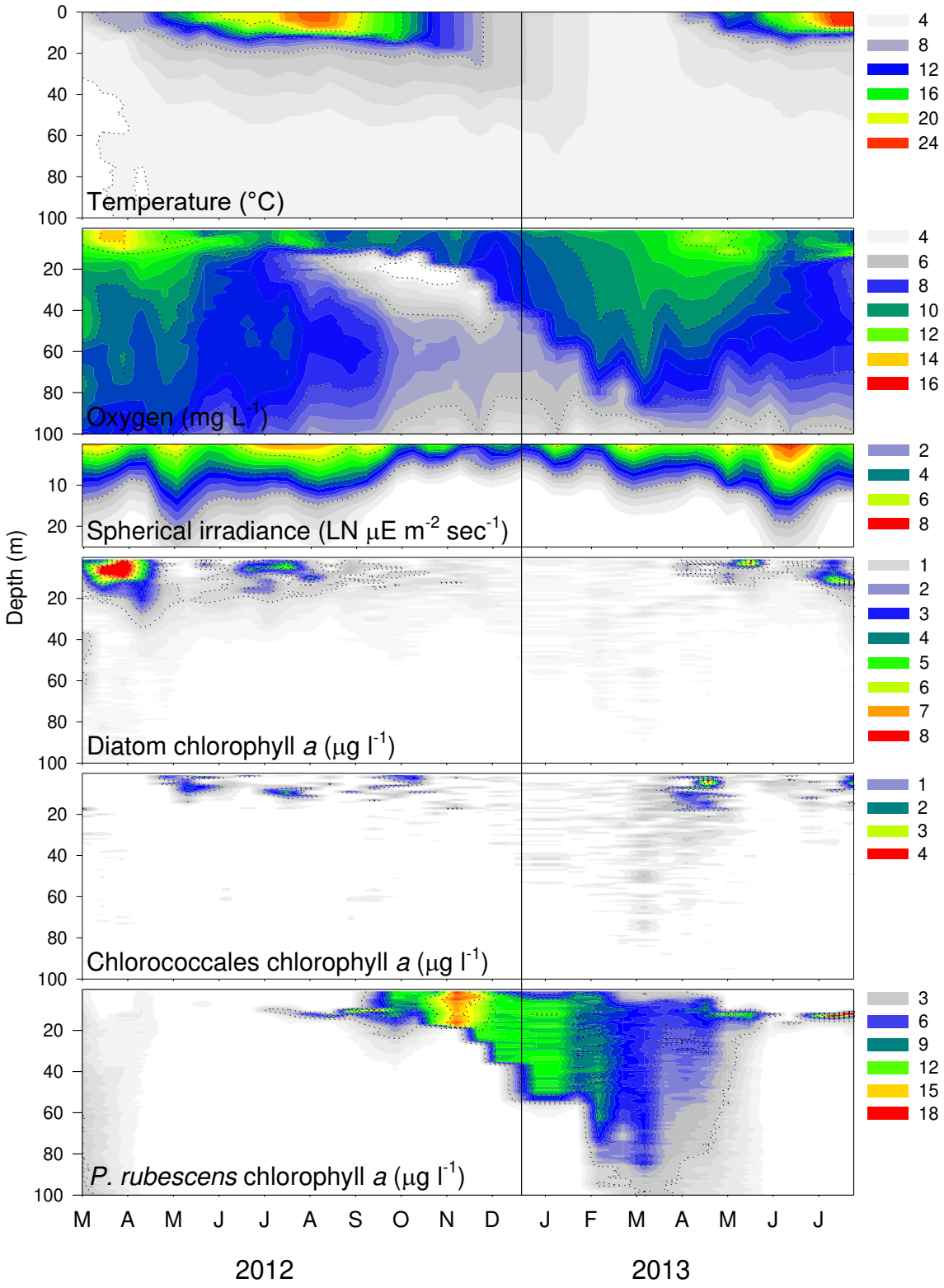


Figure S11: Redundancy analysis of environmental parameters explaining the variability in cell numbers of microbes affiliated to all '*Ca*. Nanopelagicales', '*Ca*. Nanopelagicus', and '*Ca*. P. vernalis' in Lake Zurich. temp, water temperature; picocyano, abundance of picocyanobacteria; irrad, irradiation; chloro, chlorophyll *a* associated with chlorophytes; diatom, chlorophyll *a* associated with diatoms; NH4, ammonium concentrations; O2, oxygen concentrations; PO4, phosphate concentrations; NO3, nitrate concentrations; depth, sampling depth.

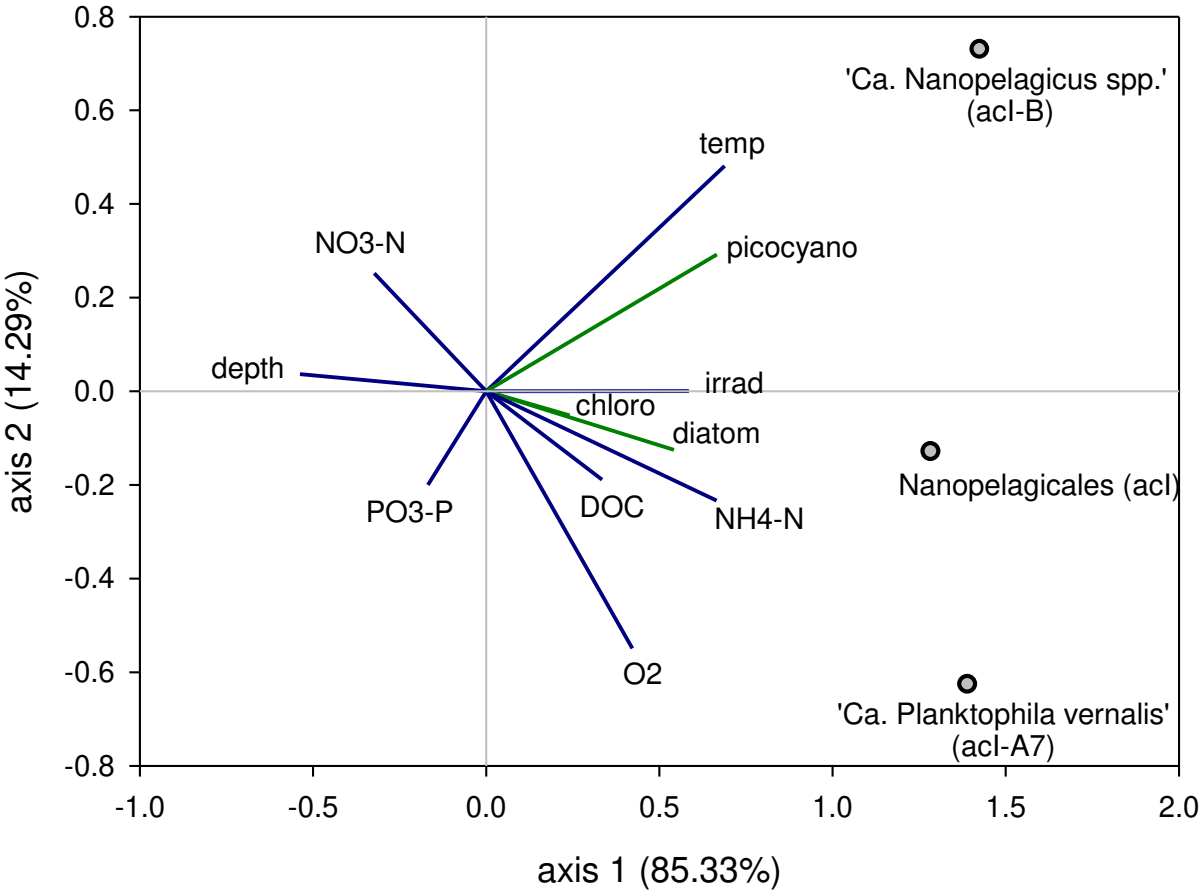


Figure S12: Metagenomic fragment recruitment of '*Ca*. Nanopelagicus' (a) and '*Ca*. Planktophila' (b) across diverse freshwater ecosystems (see Table S7 for details).

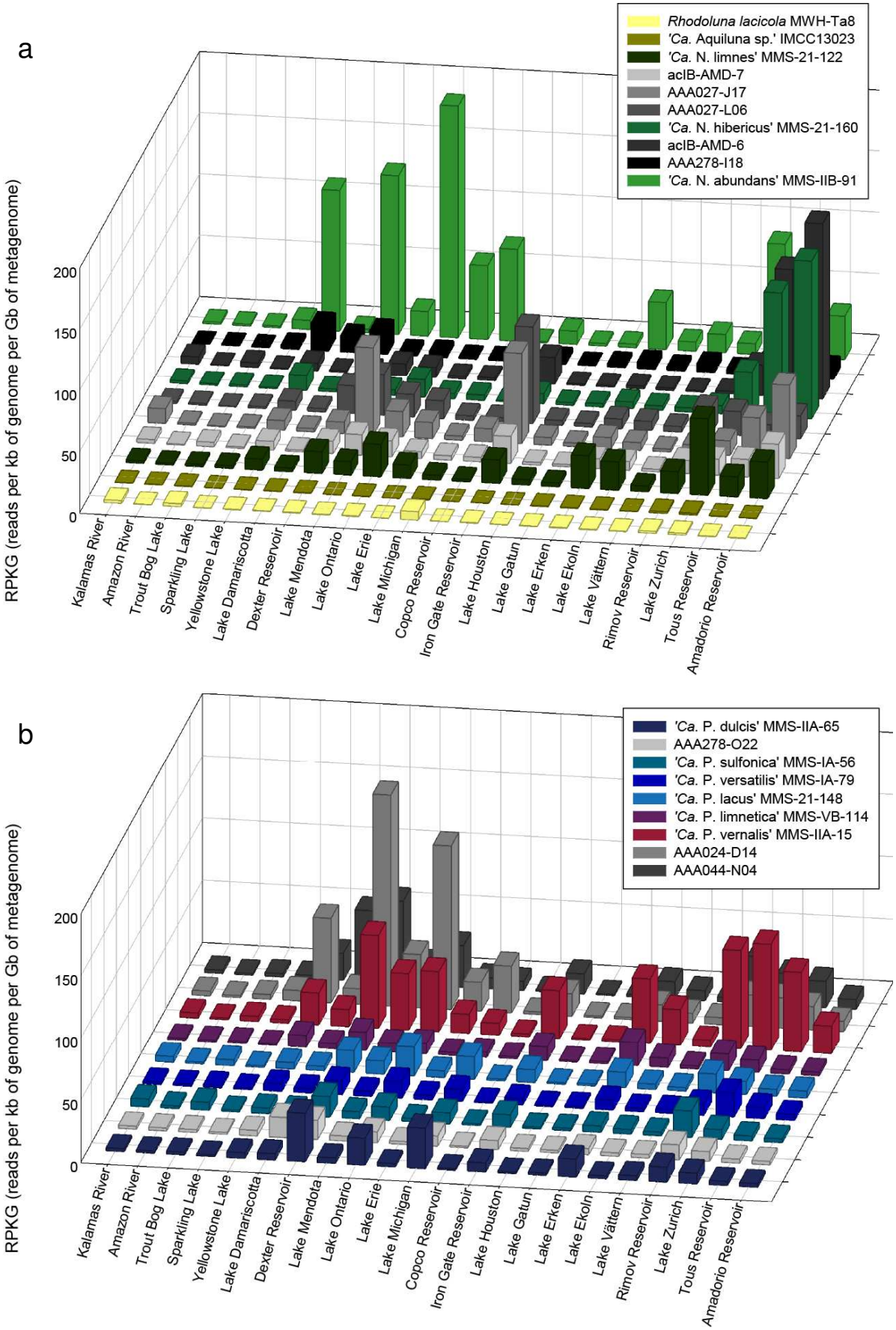


Figure S13: Recruitment plots of time-series metagenomes from Lake Mendota, USA.

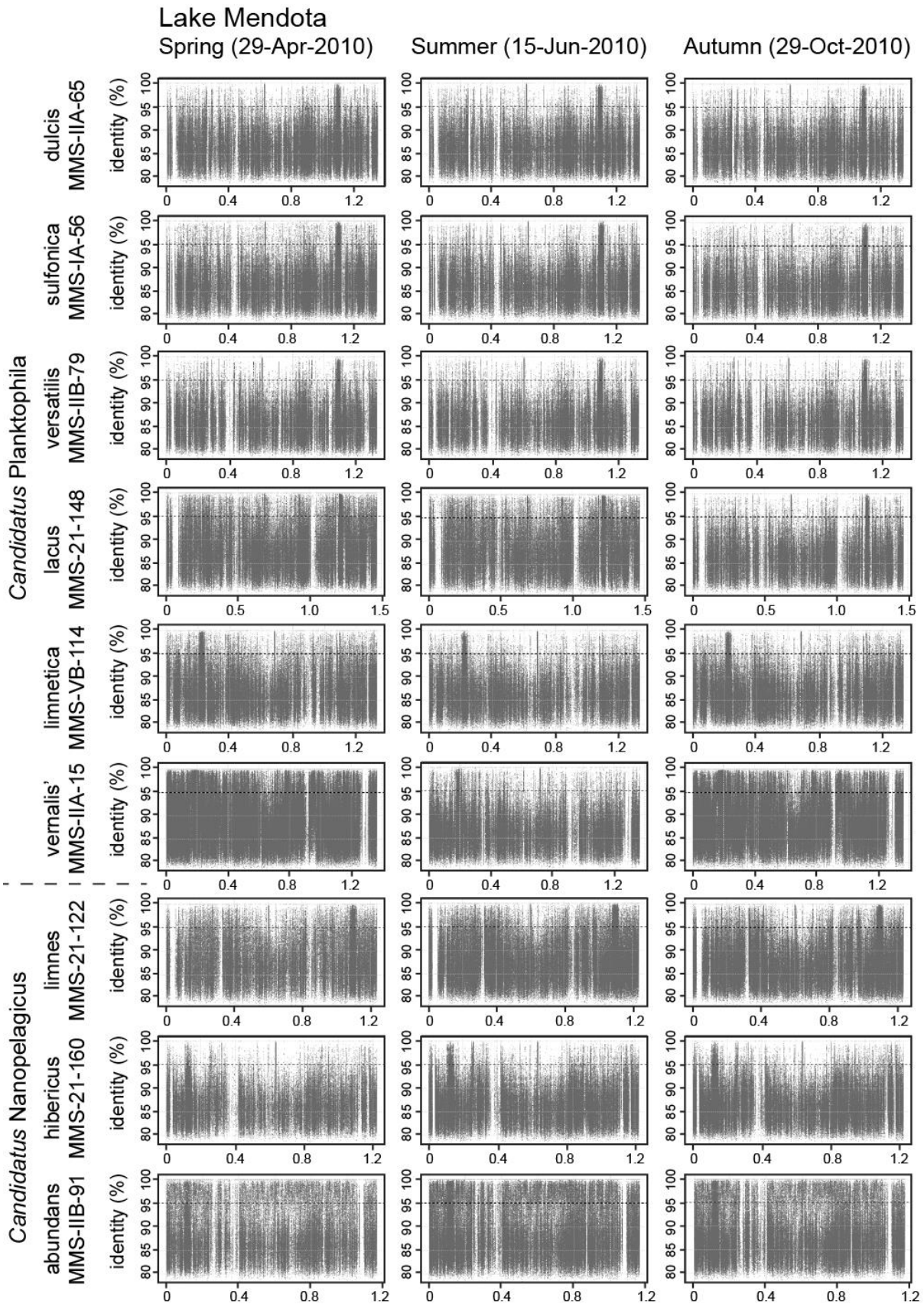


Figure S14: Whole-genome alignment of all 13 '*Ca*. Planktophila' strains. The genomes have been linearized for simplicity and are arranged in the same order as in the phylogenomic tree in Fig. 1b. Synteny and different degrees of sequence similarity are indicated by vertical lines connecting the genomes.

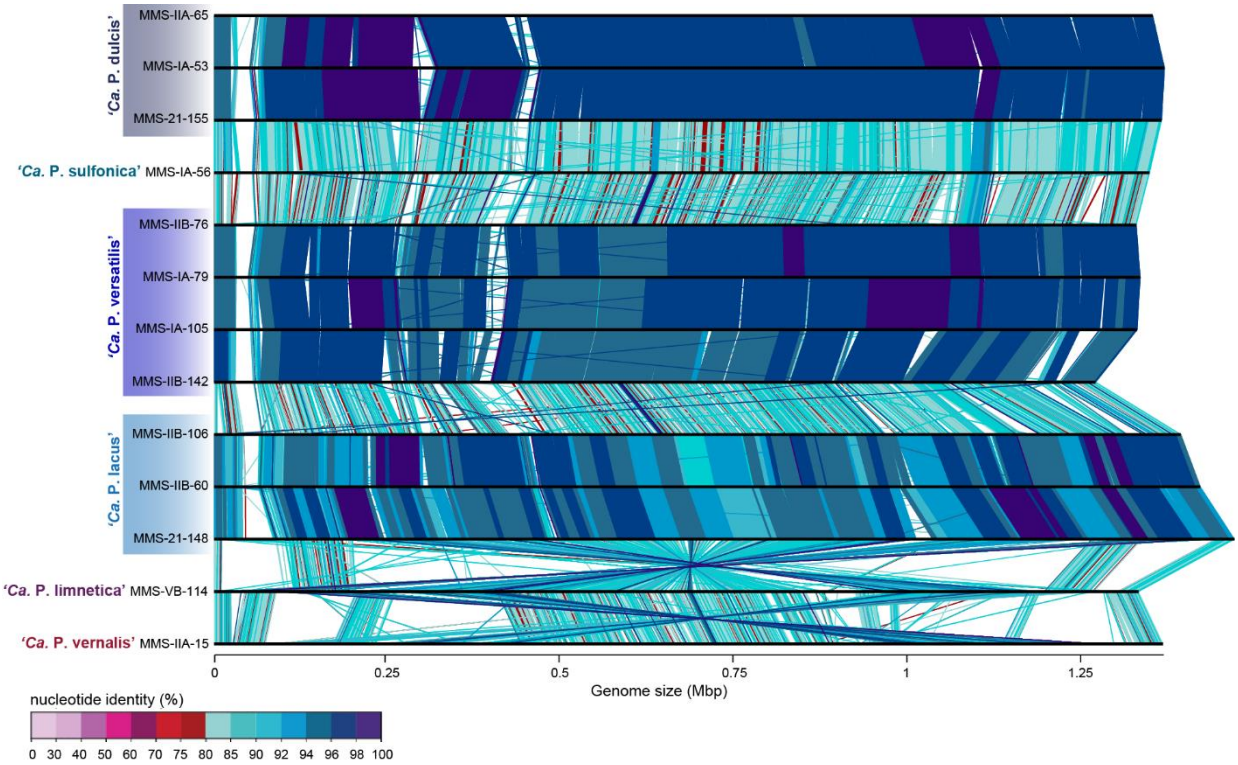


Figure S15: Whole-genome alignment of the three '*Ca*. Nanopelagicus' strains. The genomes have been linearized for simplicity and are arranged in the same order as in the phylogenomic tree in Fig. 1b. Synteny and different degrees of sequence similarity are indicated by vertical lines connecting the genomes. rRNA operons in the individual genomes are displayed as red arrows and tRNAs as short vertical lines. Genomic islands (GI) have been marked in different colours and numbered (see Table S10 for genes encoded in each island), except for '*Ca*. N. limnes' MMS-21-122, which displayed a large inversion in the genome. Red: genes encoding mainly cell wall biosynthesis and modifications; Yellow: genes encoding mainly membrane transport and / or carbohydrate metabolism.

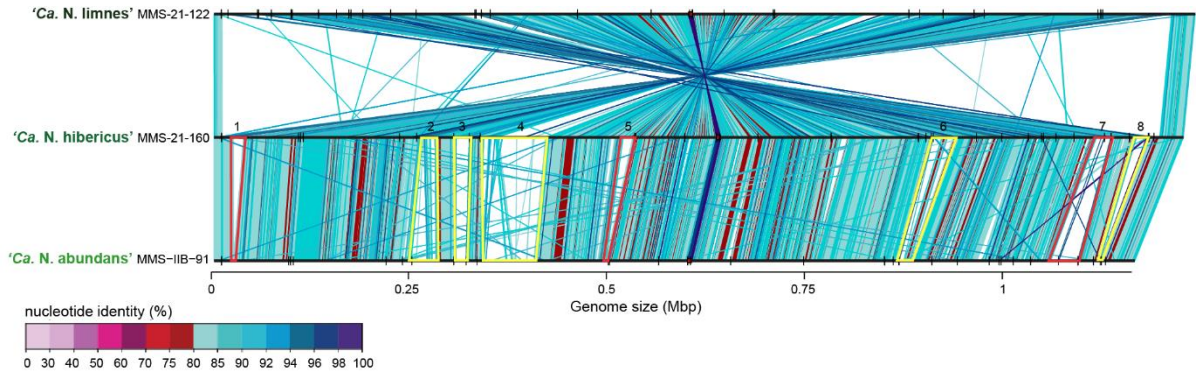


Table S1: Details of the isolation experiments conducted in Lake Zurich, Switzerland.

Date	water temp	med	# wells inoc	# cells /well	# + wells	# pure cult	# acI	label	affiliation
25/05/2012	13.5	1	168	2.0	18	17	4	MMS-21	Pdul, Plac, Nlim, Nhib
22/05/2013	13.2	2	120	1.6	20	11	4	MMS-IA	Pdul, Psul, Pvers
23/05/2013	13.2	3	168	2.0	128	95	2	MMS-IIA	Pdul, Pvern
23/05/2013	13.2	2	168	2.0	92	49	5	MMS-IIB	Pvers, Plac, Nabu
15/07/2013	20.1	4	144	1.8	51	30	1	MMS-VB	Plim
SUM			**768**		**309**	**202**	**16**		

Abbreviations: water temp, water temperature; med, medium used for dilution-to-extinction isolation; # wells inoc, number of wells inoculated; # cells/well, number of cells inoculated per well; # + wells, number of wells with dense cultures; # pure cult, number of pure cultures; # acI, number of cultures affiliated to acI Actinobacteria; label, labeling suffix for pure cultures; Nhib, '*Ca.* Nanopelagicus hibericus'; Nlim, '*Ca.* Nanopelagicus limnes'; Nabu, '*Ca.* Nanopelagicus abundans'; Pdul, '*Ca.* Planktophila dulcis'; Plim, '*Ca.* Planktophila limnetica'; Pvern, '*Ca.* Planktophila vernalis; Psul, '*Ca.* Planktophila sulfonica', Pvers, '*Ca.* Planktophila versatilis', Plac, '*Ca.* Planktophila lacus'.

Media:

1: filtered and autoclaved lake water (LW) + 10x MC#2 + vitamins (V) + amino acids (AA)
2: LW + MC#2 + MC#3 + inorganic basal medium (IBM) + stock4 + V + 10xAA
3: LW + pyruvate (50 µM) + urea (0.5 µM) + MC#4 + V + 10xAA
4: LW + pyruvate (50 µM) + urea (0.5 µM) + V + 10xAA

Minimal carbon medium MC#2: 1 µM NH_4Cl, 0.1 µM K_2HPO_4, 55.5 µM D-glucose, 66.6 µM D-ribose, 217.2 µM formate, 217.2 µM ethanol, 84.7 µM succinate, 131.5 µM glycolate, 108.6 µM glycerol, 45.2 µM N-acetylglucosamine

MC#3: 62.1 µM putrescine, 39.3 µM spermidine, 66.6 µM D-xylose, 66.6 µM arabinose

MC#4: 30 µM NH_4Cl, 50 µM oxaloacetate, 50 µM taurine, 1 µM betaine, 40 µM $CaCl_2$

Amino acids (AA): 0.5 µM isoleucine, leucine, lysine, methionine, alanine, phenylalanine, threonine, tryptophane, valine, arginine, histidine, asparagine, aspartate, cysteine, proline, serine, tyrosine, 2 µM glutamine, glutamate, glycine

Vitamins (V): 0.593 µM thiamine, 0.08 µM niacin, 0.000074 µM cobalamine, 0.005 µM para-amino benzoic acid, 0.074 µM pyridoxine, 0.081 µM pantothenic acid, 0.004 µM biotin, 0.004 µM folic acid, 0.555 µM myo-inositol

Inorganic basal medium (IBM, Hahn et al. 2003): 304 µM $MgSO_4$, 182 µM $Ca(NO_3)_2$, 190 µM $NaHCO_3$, 20 µM KCl, 16 µM K_2HP_4, 17 µM Na_2EDTA-Fe, 0.1 ml TES

TES (trace element solution): 12 µM $FeCl_3$, 16 µM H_3BO_3, 1 µM $MnCl_2$, 0.1 µM $ZnSO_4$, 0.04 µM $CuSO_4$, 0.04 µM $CoCl_2$, 0.03 µM $NaMoO_4$, 0.4 µM $NiCl_2$

Reference:

Hahn M, Lünsdorf H, Wu Q, Schauer M, Höfle M, Boenigk J *et al* (2003). Isolation of novel ultramicrobacteria classified as *Actinobacteria* from five freshwater habitats in Europe and Asia. *Appl Environ Microbiol* **69:** 1442-1451.

Table S2: Preprocessing and assembly parameters:

sample	strain	library prep	MiSeq reagents	read filtering*	assembler
MMS-IIA-65	'Ca. Planktophila dulcis'	TruSeq PCR-free	v2, 500 cycle	t, p, k	SPAdes-3.7.0
MMS-IA-53	'Ca. Planktophila dulcis'	TruSeq PCR-free	v2, 500 cycle	t, p, k	SPAdes-3.6.2
MMS-21-155	'Ca. Planktophila dulcis'	TruSeq PCR-free	v2, 500 cycle	t, p, k	a5_miseq_linux_20150522
MMS-IA-56	'Ca. Planktophila sulfonica'	TruSeq PCR-free	v2, 500 cycle	t, p, k	SPAdes-3.6.2
MMS-IIB-76	'Ca. Planktophila versatilis'	TruSeq PCR-free	v2, 500 cycle	t, p, k	SPAdes-3.6.2
MMS-IA-79	'Ca. Planktophila versatilis'	TruSeq PCR-free	v2, 500 cycle	t, p, k	SPAdes-3.7.0
MMS-IA-105	'Ca. Planktophila versatilis'	TruSeq PCR-free	v2, 300 cycle	t, p, k	a5_miseq_linux_20140604
MMS-IIB-142	'Ca. Planktophila versatilis'	TruSeq PCR-free	v2, 500 cycle	t, p, k	a5_miseq_linux_20150522
MMS-IIB-106	'Ca. Planktophila lacus'	TruSeq PCR-free	v2, 500 cycle	t, p, k	SPAdes-3.6.2
MMS-IIB-60	'Ca. Planktophila lacus'	TruSeq PCR-free	v2, 500 cycle	t, p, k	SPAdes-3.6.2
MMS-21-148	'Ca. Planktophila lacus'	TruSeq PCR-free	v2, 500 cycle	-	SPAdes-3.7.0
MMS-VB-114	'Ca. Planktophila limnetica'	TruSeq PCR-free	v2, 500 cycle	t, p	SPAdes-3.7.0
MMS-IIA-15	'Ca. Planktophila vernalis'	TruSeq PCR-free	v2, 500 cycle	t, p, k	SPAdes-3.6.2
MMS-21-122	'Ca. Nanopelagicus limnes'	TruSeq PCR-free	v2, 500 cycle	-	SPAdes-3.7.0
MMS-21-160	'Ca. Nanopelagicus hibericus'	TruSeq PCR-free	v2, 500 cycle	-	a5_miseq_linux_20150522
MMS-IIB-91	'Ca. Nanopelagicus abundans'	TruSeq PCR-free	v2, 300 cycle	t, p, k	a5_miseq_linux_20140604

*__t:__ trimmomatic-0.32.jar (MMS-IA-105, MMS-IA-105), trimmomatic-0.35.jar (others), **parameters:** PE ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10; __p:__ prinseq-lite-0.20.4, **parameters:** -out_format 3 -range_len 60-160 / 60-260 (MMS-IA-105, MMS-IA-105 / others) -range_gc 30-80 -min_qual_mean 20 -ns_max_n 2 -derep 14 -derep_min 2 -noniupac -trim_qual_right 18 -trim_qual_left 18 -lc_method entropy -lc_threshold 15 -trim_tail_left 3 -trim_tail_right 3 -trim_right 1; __k:__ kmernorm-1.0.5, **parameters:** -k 16 -c 2 -t 30.

The REPLI-g single cell kit (Qiagen, Venlo, Netherlands) was used for multiple displacement amplification (MDA). All pre-amplification steps were performed in a particle free environment dedicated to MDA. Fresh PCR-clean pipet tips were used for each MDA session and reaction tubes and PCR plates were UV-treated before usage. MDA was conducted according to the manufacturer's protocol with the following modifications: Reaction volumes were reduced to 12.5 µl and lysates for 6 to 8 replicate MDA reactions were produced in a 0.5 ml reaction tube (8 µl sample containing 2000 to 20000 cells, 6 µl each of the reagents D2 and stop) and subsequently distributed among wells of 96 well plates each containing 10 µl MDA reaction mix and SYBR I green (0.2x final concentration). Each MDA-reaction contained 2'000 - 20'000 cells.

Table S3: Details of the sequenced strains of planktonic '*Ca*. Nanopelagicales'.

taxonomy	strain	isolation date	genome size (Mbp)	GC content (%)	# CDS	# tRNA	spacer length (bp)	coding density (%)	# σ factors
'*Ca*. Planktophila dulcis'	MMS-IIA-65	23.05.2013	1.35	48.0	1344	40	12	95.7	3
'*Ca*. Planktophila dulcis'	MMS-IA-53	22.05.2013	1.37	48.0	1356	40	12	95.7	3
'*Ca*. Planktophila dulcis'	MMS-21-155	25.05.2012	1.36	47.9	1361	40	11	95.7	3
'*Ca*. Planktophila sulfonica'	MMS-IA-56	22.05.2013	1.34	48.6	1336	37	12	96.0	4
'*Ca*. Planktophila versatilis'	MMS-IIB-76	23.05.2013	1.33	48.2	1318	40	14	95.3	4
'*Ca*. Planktophila versatilis'	MMS-IA-79	22.05.2013	1.33	48.3	1327	38	14	95.2	4
'*Ca*. Planktophila versatilis'	MMS-IA-105	22.05.2013	1.33	48.2	1329	40	14	95.4	4
'*Ca*. Planktophila versatilis'	MMS-IIB-142	23.05.2013	1.27	48.3	1258	39	14	95.5	4
'*Ca*. Planktophila lacus'	MMS-IIB-106	23.05.2013	1.38	47.8	1368	40	15	95.6	4
'*Ca*. Planktophila lacus'	MMS-IIB-60	23.05.2013	1.41	47.8	1389	40	16	95.5	5
'*Ca*. Planktophila lacus'	MMS-21-148	25.05.2012	1.46	47.5	1438	41	15	95.3	4
'*Ca*. Planktophila limnetica'	MMS-VB-114	15.07.2013	1.33	45.0	1333	41	10	96.0	4
'*Ca*. Planktophila vernalis'	MMS-IIA-15	23.05.2013	1.36	45.7	1355	39	10	95.7	4
'*Ca*. Nanopelagicus limnes'	MMS-21-122	25.05.2012	1.24	41.5	1216	38	11	95.7	3
'*Ca*. Nanopelagicus hibericus'	MMS-21-160	25.05.2012	1.22	42.4	1211	38	13	95.4	4
'*Ca*. Nanopelagicus abundans'	MMS-IIB-91	23.05.2013	1.16	40.2	1150	39	14	95.3	2

Table S4: Cell sizes of different strains and *in-situ* in Lake Zurich. Asterisks indicate values based on <10 measured cells and should be treated with caution.

strain	length (µm)	s.d.	width (µm)	s.d.	volume ($µm^3$)	s.d.	CC (fg C $cell^{-1}$)	s.d.	n
MMS-IIA-65	0.44	±0.09	0.26	±0.03	0.019	±0.007	7.1	±2.3	71
MMS-IA-53	0.44	±0.14	0.28	±0.04	0.023	±0.012	8.5	±3.7	35
MMS-21-155	0.43	±0.12	0.25	±0.02	0.018	±0.007	6.8	±2.4	77
MMS-IA-56	0.50	±0.14	0.28	±0.06	0.026	±0.021	9.3	±6.4	23
MMS-IIB-76	0.49	±0.12	0.25	±0.03	0.021	±0.009	7.6	±2.8	18
MMS-IA-79	0.45	±0.10	0.27	±0.04	0.022	±0.010	8.1	±3.1	51
MMS-IA-105	0.47	±0.15	0.29	±0.05	0.027	±0.014	9.6	±4.4	64
MMS-IIB-142	0.50	±0.47	0.28	±0.29	0.026	±0.024	9.3	±8.8	23
MMS-IIB-106*	0.50	±0.14	0.30	±0.04	0.029	±0.015	10.3	±4.5	10
MMS-IIB-60	0.47	±0.08	0.30	±0.07	0.028	±0.013	10.0	±4.0	18
MMS-21-148	0.41	±0.10	0.30	±0.06	0.025	±0.014	8.8	±4.4	128
MMS-VB-114	0.38	±0.10	0.25	±0.06	0.016	±0.009	6.0	±3.0	52
MMS-IIA-15	0.43	±0.12	0.26	±0.05	0.020	±0.012	7.3	±3.9	52
MMS-21-122	0.45	±0.09	0.25	±0.03	0.018	±0.009	6.9	±2.9	85
MMS-21-160	0.33	±0.07	0.24	±0.03	0.012	±0.005	4.8	±1.7	49
MMS-IIB-91*	0.46	±0.47	0.26	±0.26	0.020	±0.020	7.5	±7.4	5

probe (sampling date)	length (µm)	s.d.	width (µm)	s.d.	volume ($µm^3$)	s.d.	CC (fg C $cell^{-1}$)	s.d.	n
Npel-2669 (18/07/2012)	0.35	±0.09	0.25	±0.04	0.014	±0.008	5.4	±2.6	97
Pver-1420 (18/07/2012)	0.37	±0.12	0.24	±0.04	0.015	±0.011	5.7	±3.4	91
Pver-1420 (15/05/2013)	0.35	±0.09	0.23	±0.03	0.012	±0.005	4.8	±1.8	87

s.d., standard deviation; CC, carbon content; n, number of cells measured

Table S5: Cell sizes and genomic details of genome-streamlined microbes.

organism	habitat	cell volume (μm^3)	source	genome size (Mbp)	GC (%)	coding density (%)	rho.	ref.
'*Ca*. Nanopelagicus'	fw	0.012-0.020	cultures	1.16-1.24	40.2-42.4	95.3-95.7	x	this study
'*Ca*. Planktophila'	fw	0.016-0.029	cultures	1.27-1.46	45.0-48.6	95.2-96.0	x	this study
'*Ca*. Nanopelagicales'	fw	N.A.	SAGs	1.10-1.66	41.4-47.6	95.8-96.3	x	1, 2
'*Ca*. Nanopelagicus'	fw	N.A.	MAGs	1.38-2.2	40.4-42.1	96.1-96.4	x	3
'*Ca*. Actinomarina minuta'	mar	0.013	MAG	0.8-1.03	33.4		x	4
'*Ca*. Pelagibacter' (SAR11)	mar	0.025-0.045	cultures	1.24-1.48	28.6-32.3	93.6-97	x	5, 6, 7
LD12	fw	0.017	SAGs	1.03-1.39	29-30	95.5-96.5	x	8, 9
'*Ca*. Methylopumilus planktonicus'	fw	0.030-0.075	cultures	1.28-1.35	37.0-37.3	94.9-95.3	x	10, unpubl.
OM43	mar	0.023	cultures	1.30-1.37	35.4-37.9	95.0-96.7	x	11
Rhodoluna lacicola	fw	0.053	cultures	1.43	51.5	93.7	x	12
'*Ca*. Aquiluna sp.'	fw	N.A.	cultures	1.36	51.7	93.5	x	13

fw, freshwater; mar, marine; rho., rhodopsins; ref., references; N.A., data not available

References:

[1] Ghylin TW, Garcia SL, Moya F, Oyserman BO, et al. (2014) Comparative single-cell genomics reveals potential ecological niches for the freshwater acI Actinobacteria lineage. *ISME Journal* **8**:2503-2516

[2] Garcia SL, McMahon KD, Martinez-Garcia M, Srivastava A, et al. (2013) Metabolic potential of a single cell belonging to one of the most abundant lineages in freshwater bacterioplankton. *ISME Journal* **7**:137-147

[3] Ghai R, Mizuno CM, Picazo A, Camacho A, Rodriguez-Valera F (2014) Key roles for freshwater Actinobacteria revealed by deep metagenomic sequencing. *Molecular Ecology* **23**:6073-6090

[4] Ghai R, Mizuno CM, Picazo A, Camacho A, Rodriguez-Valera F (2013) Metagenomics uncovers a new group of low GC and ultra-small marine Actinobacteria. *Scientific Reports* **3**

[5] Nicastro D, Schwartz C, Pierson J, Cho J-C, et al. (2006) Three-dimensional structure of the tiny bacterium *Pelagibacter ubique* studied by cryo-electron tomography. *Microscopy and Microanalysis* **12**:180-181

[6] Zhao X, Schwartz C, Pierson J, Giovannoni SJ, et al. (2016) Three-dimensional structure of the ultra-oligotrophic marine bacterium Pelagibacter. *Applied and Environmental Microbiology*

[7] Grote J, Thrash JC, Huggett MJ, Landry ZC, et al. (2012) Streamlining and core genome conservation among highly divergent members of the SAR11 clade. *mBio* **3**

[8] Salcher MM, Pernthaler J, Posch T (2011) Seasonal bloom dynamics and ecophysiology of the freshwater sister clade of SAR11 bacteria 'that rule the waves' (LD12). *ISME Journal* **5**:1242-1252

[9] Eiler A, Mondav R, Sinclair L, Fernandez-Vidal L, et al. (2016) Tuning fresh: radiation through rewiring of central metabolism in streamlined bacteria. *ISME Journal* **10**:1902-1914

[10] Salcher MM, Neuenschwander SM, Posch T, Pernthaler J (2015) The ecology of pelagic freshwater methylotrophs assessed by a high-resolution monitoring and isolation campaign. *ISME Journal* **9**:2442-2453

[11] Jimenez-Infante F, Ngugi DK, Vinu M, Alam I, et al. (2016) Comprehensive genomic analyses of the OM43 clade, including a novel species from the Red Sea, indicate ecotype differentiation among marine methylotrophs. *Applied and Environmental Microbiology* **82**:1215-1226

[12] Hahn MW, Schmidt J, Taipale SJ, Doolittle WF, Koll U (2014) Rhodoluna lacicola gen. nov., sp. nov., a planktonic freshwater bacterium with stream-lined genome. *International Journal of Systematic and Evolutionary Microbiology* **64**:3254-3263

[13] Kang I, Lee K, Yang S-J, Choi A, et al. (2012) Genome sequence of "*Candidatus* Aquiluna" sp. strain IMCC13023, a marine member of the Actinobacteria isolated from an Arctic fjord. *Journal of Bacteriology* **194**:3550-3551

Table S6: Details of the applied oligonucleotide probes

probe	specificity	sequence (5' to 3')	target (rRNA, 5' position)	FA%
Npel-23S-2669	'*Ca*. Nanopelagicus' (acI-B1)	ACA AGA GGT TCG TCC GTC C	23S, 2669	60
Npel-C1	competitor 1 for Npel-23S-2669	ACY AGA GGT TCG TCC GTC C		
Npel-C2	competitor 2 for Npel-23S-2669	ACA AGA GGT TCG TCC ATC C		
Npel-C3	competitor 3 for Npel-23S-2669	ACY AGA GGT TCG TCC ATC C		
Npel-H1	helper 1 for Npel-23S-2669	CGG TCC TCT CGT ACT AGG GAC AGC		
Npel-H2	helper 2 for Npel-23S-2669	GTG CTY CTG GCG RAA CAA CCG ACA C		
Npel-H3	helper 3 for Npel-23S-2669	CYT TCC RAA CGT TGC WAA TCG GCC		
Pver-23S-1420	'*Ca*. Planktophila vernalis' (acI-A7)	AAC TAC TAC CAC ACC GGT TCG	23S, 1420	55
Pver-C1	competitor 1 for Pver-23S-1420	AAC TAC TAC CAC ACC GGT TCA		
Pver-C2	competitor 2 for Pver-23S-1420	AAC TAC TAC CAC ACC GGG TCG		
Pver-C3	competitor 3 for Pver-23S-1420	AAC TAC TAC AAC ACC GGT TCG		
Pver-C4	competitor 4 for Pver-23S-1420	AAC TAC TAC AAC ACC GGG TCA		
Pver-H1	helper 1 for Pver-23S-1420	CAT TAG TGG RTT CGT YAT GGG CGA ATT A		
Pver-H2	helper 2 for Pver-23S-1420	AGC CAT CCA CCC ACG CRG CTT		
Pver-H3	helper 3 for Pver-23S-1420	CTG TGT CAC ACC ATT GCT T		
AcI-852[1]	'*Ca*. Nanopelagicales' (acI lineage)	AAT GCG TTA GCT GCG TCG CA	16S, 852	55
AcI-852-H1[1]	Helper for AcI-852	AAA CCG TGG AAG GTY CSC ACA ACT AG		
AcI-852-H2[1]	Helper for AcI-852	TCC CCA GGC GGG GCR CTT		

Abbreviation: FA%, formamide concentration required for the CARD-FISH hybridization buffer.

Reference:

[1] Warnecke F, Sommaruga R, Sekar R, Hofer JS, Pernthaler J (2005). Abundances, identity, and growth state of *Actinobacteria* in mountain lakes of different UV transparency. *Appl Environ Microbiol* **71:** 5551-5559